



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

ROZPOZNÁVÁNÍ VÝRAZU TVÁŘE

FACIAL EXPRESSION RECOGNITION

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Markéta Vránová

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Martin Mézl

BRNO 2016

Diplomová práce

magisterský navazující studijní obor **Biomedicínské inženýrství a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Bc. Markéta Vránová

ID: 147524

Ročník: 2

Akademický rok: 2015/16

NÁZEV TÉMATU:

Rozpoznávání výrazu tváře

POKYNY PRO VYPRACOVÁNÍ:

1) Seznamte se s metodami detekce obličeje v obrazových datech. 2) Proveďte rešerši dostupných metod pro rozpoznání výrazu tváře ve statickém obraze nebo videu. 3) Realizujte základní algoritmy pro detekci obličeje v obraze. Metodu otestujte na vhodných datech. 4) Realizujte algoritmy pro základní rozpoznávání výrazu tváře a proveďte testování na referenčních snímcích. 5) Zrealizujte snímání databáze alespoň deseti dobrovolníků. Na takto získaných datech proveďte testování algoritmů. 6) Proveďte diskuzi dosažených výsledků.

DOPORUČENÁ LITERATURA:

[1] YANG, M. H., KRIEGMAN D. J., AHUJA, N. Detecting faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002, vol. 24, no. 1, p. 34-58.

[2] MISHRA, B. et al. Facial expression recognition using feature based techniques and model based techniques: A survey. 2015 2nd International Conference on Electronics and Communication Systems (ICECS). IEEE, 2015, : 589-594.

Termín zadání: 8.2.2016

Termín odevzdání: 20.5.2016

Vedoucí práce: Ing. Martin Mézl

Konzultant diplomové práce:

prof. Ing. Ivo Provazník, Ph.D., předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato práce se zabývá automatickým rozpoznáním výrazů tváře v barevných obrazech. Nejprve jsou podle barvy kůže detekovány obličej v obrazech. Detekce využívá barevných prostorů RGB, HSV a YCbCr. Následuje automatický ořez oblasti obličeje provedený rozměření okrajů obličeje podle pozic základních rysů (oči, nos a ústa). Z oblasti obličeje je pomocí banky filtrů získán vektor příznaků. Práce uvádí dvě různé realizace Gaborových filtrů a je doplněna o návrh vlastní banky filtrů. Získaný vektor příznaků vstupuje do neuronové sítě, která obraz klasifikuje. Neuronová síť byla naučena na množině obrazů z AR databáze určené pro rozpoznávání výrazů. Výstupem je přiřazení výrazu obličeje v obraze k některému z předem určených výrazů. Práce uvádí testování parametrů neuronových sítí a dále prezentuje a diskutuje dosažené výsledky.

KLÍČOVÁ SLOVA

Rozpoznání výrazů tváře, výraz tváře, neuronová síť, klasifikace, Gaborovy filtry, detekce obličeje, barva kůže, detekce podle barvy, barevný obraz, barevné prostory

ABSTRACT

This project deals with automatic recognition of facial expression in colour pictures. At first, the colour-based face detection is accomplished, three colour spaces are used: RGB, HSV and YCbCr. As next, the pictures are automatically cropped so that only the face region is present. It is accomplished by computing the borders of the face region based on knowledge of position of eyes, nose and mouth. From the face region, the feature vector is obtained using a bank of Gabor filters. The project introduces two different kinds of Gabor filters and proposes a new bank of filters. The feature vector is used as an input to the neural network. The neural network was trained on a set of pictures from AR database created for facial expression recognition. The output of the network is the facial expression the input picture was assigned to. This project mentions the testing for different settings of the neural network and presents and discuss the recognition results of the network.

KEYWORDS

Facial Expression detection, facial expression, neural network, classification, Gabor filters, face detection, skin colour, colour-based detection, colour pictures, colour spaces

VRÁNOVÁ, M. *Rozpoznání výrazu tváře*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2016. 72 s. Diplomová práce. Vedoucí práce: Ing. Martin Mézl.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma Rozpoznání výrazu tváře jsem vypracovala samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne 18. 5. 2016

.....

(podpis autora)

PODĚKOVÁNÍ

Děkuji svému vedoucímu diplomové práce Ing. Martinu Mězlovi za ochotnou pomoc a dobré připomínky a vedení mé diplomové práce.

V Brně dne 18. 5. 2016

.....

(podpis autora)

OBSAH

Seznam obrázků	vii
Seznam tabulek	ix
Úvod	1
1 Digitální reprezentace obrazu	2
1.1 Barevné prostory	2
1.1.1 Barevný prostor RGB (RGBA).....	2
1.1.2 CMY (CMYK).....	3
1.1.3 HSV prostor	4
1.1.4 HSL prostor.....	5
1.1.5 YCbCr prostor.....	6
1.1.6 YIQ	7
1.1.7 YUV	8
1.1.8 CIE L*a*b.....	8
1.1.9 Další barevné prostory	9
1.2 Barevné formáty	10
1.2.1 Kompresní metody.....	10
1.2.2 Používané barevné formáty	11
2 Detekce obličeje v obraze	13
2.1 Příznakové metody (Feature-based Approach).....	14
2.1.1 Metody využívající základní charakteristiky pixelu nebo jeho malého okolí (Low-level Analysis)	15
2.1.2 Metody hledání rysů obličeje (Feature Analysis)	16
2.1.3 Metoda aktivních kontur	18

2.2	Metody založené na obrazové informaci (Image-based Approach)	19
2.2.1	Lineární podprostorové metody	20
2.2.2	Neuronové sítě	20
2.2.3	Statistické metody	21
2.2.4	Viola-Jones algoritmus	21
2.2.5	Skrytý Markovský model.....	23
2.3	Morfologické operace	23
2.3.1	Eroze	24
2.3.2	Dilatace	24
3	Metody rozpoznávání výrazu obličeje	25
3.1	Metody extrakce příznaků	25
3.1.1	Metody s geometrickým přístupem	25
3.1.2	Metody založené na šabloně	26
3.1.3	Metody založené na jasové informaci barevného obrazu	26
3.1.4	Metody založené na vzhledu (Appearance Based Approach)	26
3.1.5	Jiné možné rozdělení metod extrakce příznaků	27
3.2	Klasifikátory výrazů obličeje.....	28
3.2.1	Skryté Markovské modely (HMM)	28
3.2.2	Neuronová síť	28
3.2.3	Support Vector Machine (SVM).....	30
3.2.4	AdaBoost	30
3.2.5	Algoritmus nejbližších sousedů (K-Nearest-Neighbor, KNN).....	30
3.2.6	Eigenface (eigenpicture, eigenvector)	30
3.2.7	Fisherfaces	31
3.2.8	Bayesovská klasifikace (Bayesian classification).....	31
3.2.9	Metody redukující vektor příznaků.....	31

4	Detekce obličeje a obličejových rysů	33
4.1	Detekce obličeje podle barvy kůže v barevných prostorech RGB, HSV a YCbCr	33
4.2	Aplikace morfologických operací.....	35
4.3	Detekce charakteristických rysů obličeje	36
4.3.1	Detekce středů očí.....	36
4.3.2	Detekce středu úst.....	38
4.3.3	Detekce špičky nosu	38
4.3.4	Automatický ořez obličeje	39
5	Rozpoznávání výrazů obličeje	44
5.1	Databáze obličejů.....	44
5.1.1	AR Face Database.....	44
5.1.2	Vlastní databáze	45
5.2	Extrakce příznaků z obrazu.....	46
5.2.1	První banka Gaborových filtrů	46
5.2.2	Druhá banka Gaborových filtrů	47
5.2.3	Vlastní návrh banky filtrů	49
5.3	Neuronová síť.....	51
5.4	Výsledky rozpoznávání výrazu tváře.....	55
6	Závěr	59
	Literatura	61
	Seznam zkratk	66
	Přílohy	68

SEZNAM OBRÁZKŮ

Obrázek 1-1 Rozmístění barev v jednotkové krychli (vlevo) a princip vytváření barevných odstínů ze tří základních barev RGB (červená, zelená a modrá) [19].....	3
Obrázek 1-2 Rozmístění barev v jednotkové krychli (vlevo) a princip vytváření barevných odstínů ze tří základních barev CMY (azurová, purpurová a žlutá) [19].....	4
Obrázek 1-3 Zobrazení kombinace tří parametrů HSV prostoru (barevný tón, sytost a jas) a barev vycházejících z těchto parametrů [19].....	5
Obrázek 1-4 Zobrazení kombinace tří parametrů HSL prostoru (barevný tón, sytost a světlost) a barev vycházejících z těchto parametrů [19].....	5
Obrázek 1-5 Zobrazení kanálů různých barevných prostorů [9]	7
Obrázek 1-6 Znázornění kanálů U a V prostoru YUV udávajících barvu pixelu [55]	8
Obrázek 1-7 Zobrazení prostoru CIE L*a*b [19].....	9
Obrázek 2-1 Znázornění posuvu okna po obraze a příklady masek Haarových filtrů (a-f) [42s]	22
Obrázek 4-1 Ukázka zdařilé detekce úst, očí a nosu. a) detekce úst v prostoru RGB, b) detekce očí první metodou, c) detekce očí druhou metodou, d) detekce očí třetí metodou, e) detekce očí, nosu a úst pomocí Viola-Jones algoritmu	39
Obrázek 4-2 Ukázka méně zdařilé detekce úst, očí a nosu. a) detekce úst v prostoru RGB, b) detekce očí první metodou, c) detekce očí druhou metodou, d) detekce očí třetí metodou, e) detekce očí, nosu a úst pomocí Viola-Jones algoritmu	39
Obrázek 4-3 Zobrazení ručního výběru oblasti obličeje pomocí funkce imrect.....	41
Obrázek 4-4 Ukázka zdařilého automatického ořezu obrazu z AR databáze a) vstupní obraz, b) binární obraz po detekci bravy kůže, c) upravený binární obraz oblasti obličeje, d) obraz po automatickém ořezu	42
Obrázek 4-5 Ukázka méně zdařilého automatického ořezu obrazu z vlastní databáze a) vstupní obraz, b) binární obraz po detekci bravy kůže, c) upravený binární obraz oblasti obličeje, d) obraz po automatickém ořezu.....	42
Obrázek 4-6 Ukázka nezdařilého automatického ořezu obrazu z vlastní databáze a) vstupní obraz, b) binární obraz po detekci bravy kůže, c) upravený binární obraz oblasti obličeje, d) obraz po automatickém ořezu.....	42

Obrázek 4-7 Ukázka srovnání manuálního a automatického ořezu obrazů obličeje, obrazy a), c) a e) - manuální ořez, obrazy b), d), f) - automatické ořezy.....	43
Obrázek 4-8 Vývojový diagram algoritmu pro vyhledání oblasti obličeje v barevném vstupním obrazu a následný ořez	43
Obrázek 5-1 Ukázka obrazů databáze AR s různými výrazy obličeje bez manuálního ořezu, a) neutrální výraz, b) smích, c) vztek, d) řev	44
Obrázek 5-2 Ukázka obrazů vlastní databáze s různými výrazy obličeje bez manuálního ořezu, a) neutrální výraz, b) úsměv, c) vztek, d) smích (s otevřenou pusou), e) překvapení.....	45
Obrázek 5-3 Vývojový diagram rozpoznávání výrazu tváře	46
Obrázek 5-4 Banka deseti Gaborových 2D filtrů podle rovnice 5.1	47
Obrázek 5-5 Příklad výstupních obrazů po konvoluci s Gaborovými filtry z první banky	47
Obrázek 5-6 Banka deseti Gaborových 2D filtrů (reálná složka) podle rovnice 5.4.....	48
Obrázek 5-7 Příklad výstupních obrazů po konvoluci s Gaborovými filtry z druhé banky	49
Obrázek 5-8 Banka deseti 2D filtrů podle rovnice 5.7	50
Obrázek 5-9 Frekvenční charakteristiky 1., 2., 3. a 5. filtru získaných podle rovnice 5.7 po aplikaci Fourierovy transformace	50
Obrázek 5-10 Výsledky po aplikaci vlastní banky filtrů	50
Obrázek 5-11 Vývojový diagram procesu rozpoznávání výrazu obličeje v obraze	55

SEZNAM TABULEK

Tabulka 4-1 Přehled podmínek pro barvu kůže pro jednotlivé kanály prostorů RGB, HSV a YCbCr [50].....	34
Tabulka 4-2 Střední hodnoty poměrů vzdáleností mezi význačnými body obličeje	40
Tabulka 5-1 Výsledky testování počtů skrytých vrstev sítě a počtů neuronů ve skrytých vrstvách pro neuronové sítě učené na vektorech příznaků získaných pomocí 1. banky filtrů.....	52
Tabulka 5-2 Výsledky testování počtů skrytých vrstev sítě a počtů neuronů ve skrytých vrstvách pro neuronové sítě učené na vektorech příznaků získaných pomocí 2. banky filtrů.....	53
Tabulka 5-3 Výsledky testování počtů skrytých vrstev sítě a počtů neuronů ve skrytých vrstvách pro neuronové sítě učené na vektorech příznaků získaných pomocí 3. banky filtrů.....	54
Tabulka 5-4 Úspěšnost klasifikace pro různé sady obrazů klasifikované sítí se 2 skrytými vrstvami (100, 20), příznaky získány pomocí první banky filtrů .	56
Tabulka 5-5 Úspěšnost klasifikace pro různé sady obrazů klasifikované sítí s jednou skrytou vrstvou se 20 neurony, příznaky získány pomocí druhé banky filtrů	56
Tabulka 5-6 Úspěšnost klasifikace pro různé sady obrazů klasifikované sítí se 2 skrytými vrstvami (100, 30), příznaky získány pomocí třetí banky filtrů ...	57

SEZNAM PŘÍLOH

A	Výsledky klasifikace výrazů pomocí neuronové sítě.....	68
A.1	Výsledky klasifikace výrazů pomocí neuronové sítě naučené na výstupech vlastní banky filtrů	68
A.2	Výsledky klasifikace výrazů pomocí neuronové sítě naučené na výstupech první banky Gaborových filtrů	69
A.3	Výsledky klasifikace výrazů pomocí neuronové sítě naučené na výstupech druhé banky Gaborových filtrů	70
B	Vlastní databáze obrazů	71
C	Soupis souborů přiložených na cd	72

ÚVOD

Detekce obličeje v obraze a rozpoznání výrazu tváře je předmětem zájmu mnoha výzkumných týmů, protože neexistuje žádný deterministický způsob, jak v obraze obličeje najít, jak najít různé rysy obličeje a jak zjistit výraz obličeje. Existují metody, které tuto problematiku dobře řeší, ale zpravidla předpokládají určité vlastnosti všech vstupních obrazů. Úkolem je tedy vytvořit algoritmus, který bude univerzální a bude schopen fungovat pro různé obrazy s různorodými obličejí (různá velikost, různý počet obličejů v obraze a jejich pozice, různá barva kůže pro různé etnické skupiny osob, různé pozadí a podobně).

Detekce obličejů v obraze je první krok v rozpoznávání výrazů obličeje a je důležitou součástí počítačového vidění a mnoha biometrických aplikací, například automatické rozpoznání obličejů osob, komunikace člověka s počítačem nebo vyhledávání obličejů v obsáhlých databázích. V této práci je detekce obličeje provedena podle barvy kůže s využitím několika barevných prostorů. Metoda neklade speciální požadavky na vstupní obrazy a je relativně jednoduchá. Výběr oblasti obličeje je doplněn o detekci hlavních rysů obličeje, které pomáhají výběr oblasti obličeje zpřesnit.

Dalším krokem je rozpoznání výrazu obličeje, které může být využito například pro zjednodušení komunikace člověka s počítačem pomocí změny výrazu nebo pro analýzu lidských emocí. Rozpoznávání výrazu probíhá v oblasti obličeje ve vstupním obraze a je realizováno pomocí neuronové sítě. Do ní vstupují příznaky obrazu získané pomocí banky několika dvourozměrných filtrů. Výstupem je rozpoznáný výraz obličeje.

V teoretické části této práce jsou popsány digitální reprezentace obrazu, metody detekce obličeje a metody rozpoznávání výrazů obličejů. V praktické části je popsán postup detekce obličeje, realizace bank filtrů, realizace neuronové sítě a testování jejích parametrů.

Pro detekci obličeje byly testovány různé podmínky určující bravu kůže v různých barvených prostorech. V oblasti barvy kůže byly detekovány hlavní rysy obličeje, pomocí kterých byl vytvořen automatický výběr oblasti obličeje. Pro výpočet vektoru příznaků obrazu byly implementovány tři banky filtrů, z toho dvě banky obsahují Gaborovy filtry a třetí banka obsahuje vlastní navržené filtry. Na vektorech příznaků získaných pomocí těchto bank filtrů byly naučeny neuronové sítě. Práce uvádí testování parametrů neuronových sítí a posouzení jejich vlivu na přesnost rozpoznávání výrazu. Neuronové sítě, které vykazují pro jednotlivé banky filtrů nejlepší výsledky, byly testovány na obrazech z databáze pro rozpoznávání výrazů a na obrazech z vlastní databáze. Práce uvádí dosažené výsledky rozpoznávání výrazů pomocí neuronové sítě.

1 DIGITÁLNÍ REPREZENTACE OBRAZU

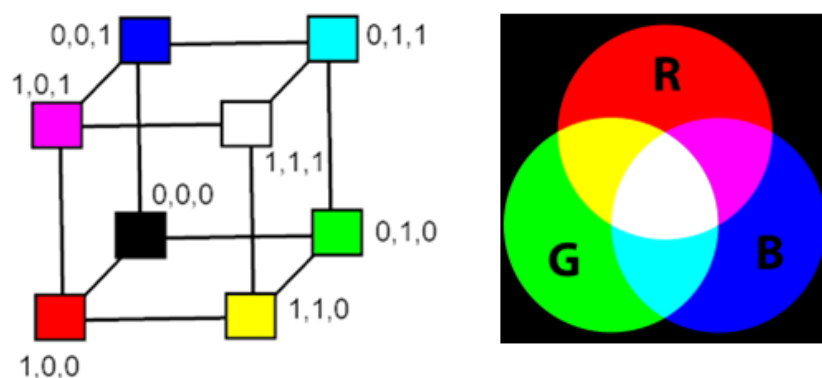
Všechny obrazy, které jsou zpracovávány při detekci obličeje nebo hledání rysů obličeje, jsou uloženy a zpracovávány v digitální podobě. Šedotónový obraz je uložen jako matice bodů, každý bod odpovídá jednomu pixelu obrazu, hodnota každého bodu udává hodnotu jasu daného pixelu. Barevné obrazy jsou uloženy pomocí několika matic, jejich počet je odvozen od použití konkrétních barevných prostorů, které jsou popsány v kapitole 1.1. Způsob uložení obrazových dat je popsán v kapitole 1.2, jsou uvedeny používané barevné formáty. [9], [25], [27]

1.1 Barevné prostory

Při práci s barevným obrazem je důležitou otázkou, jak barvy popsat nebo reprezentovat. Za tímto účelem existují různé barevné prostory, které budou v následujících podkapitolách popsány. Jejich využití závisí především na vlastnostech jednotlivých prostorů, na jejich výhodách a nevýhodách. [19], [27]

1.1.1 Barevný prostor RGB (RGBA)

Tento prostor funguje na základě aditivního sčítání barev, tři kanály odpovídají třem barvám: *R* - red - červená, *G* - green - zelená, *B* - blue - modrá. Intenzitu jednotlivých složek lze zadávat buďto celočíselně (v rozsahu 0-255) nebo hodnotami z rozsahu $\langle 0,1 \rangle$. Nulová hodnota všech kanálů dá výslednou černou barvu, naopak maximální hodnoty ve všech kanálech představují barvu bílou. Kombinování tří základních barev je znázorněno na obrázku 1-1. Výsledná barevná hloubka je 24 bitů, a je možné dostat 256^3 různých barevných odstínů (neboť každý kanál může nabývat 256 různých odstínů, prostor má tři kanály, a výsledná barva může být jakákoli kombinace hodnot tří kanálů), což je více než 16 milionů barev. Kanál *A* u prostoru RGBA představuje tak zvaný alfa kanál, který určuje průhlednost barvy, využívá se například ve formátech PNG a GIF. RGB prostor se využívá k zobrazování a ukládání dat.



Obrázek 1-1 Rozmístění barev v jednotkové krychli (vlevo) a princip vytváření barevných odstínů ze tří základních barev RGB (červená, zelená a modrá) [19]

Vyjádřením jednotlivých barev jako poměr dané barvy ku všem barvám je možné dosáhnout normalizovaných barev RGB. Normalizovaný prostor RGB umožňuje odfiltrovat vliv osvětlení a je proto využíván v detekci barvy kůže. Výpočet normalizovaných barev z RGB barev je realizovaný pomocí následujících rovnic [21]:

$$r = \frac{R}{R+G+B} \quad (1.1)$$

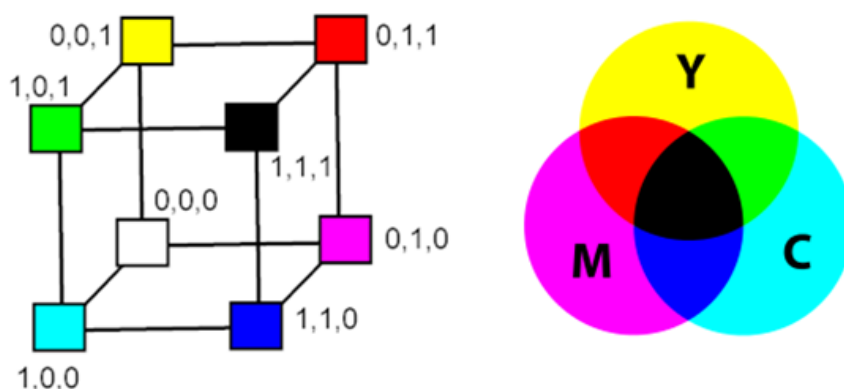
$$g = \frac{G}{R+G+B} \quad (1.2)$$

$$b = \frac{B}{R+G+B} \quad (1.3)$$

R, G, B představují kanály prostoru RGB, r, g, b označují normalizované barvy. Z těchto rovnic vyplývá, že $r+g+b = 1$ a proto k reprezentaci normalizovaných barev postačí pouze hodnoty r a g , a hodnota b může být dopočítána využitím faktu, že součet všech tří normalizovaných složek je roven jedné. [19], [21]

1.1.2 CMY (CMYK)

Tento prostor využívá subtraktivního míchání barev, tři barevné kanály jsou následující: C - cyan - azurová, M - magenta - purpurová, Y - yellow - žlutá. Nulové hodnoty všech kanálů odpovídají barvě bílé, maximální hodnoty ve všech kanálech odpovídají černé barvě, znázorněno je to na obrázku 1-2. V polygrafii se používá prostor CMYK, kdy je navíc přiřazen kanál K - black - černá.



Obrázek 1-2 Rozmístění barev v jednotkové krychli (vlevo) a princip vytváření barevných odstínů ze tří základních barev CMY (azurová, purpurová a žlutá) [19]

Nevýhodou obou výše zmíněných prostorů je, že si člověk na základě znalostí tří nebo čtyř hodnot jednotlivých kanálů těžko dokáže představit výslednou barvu.

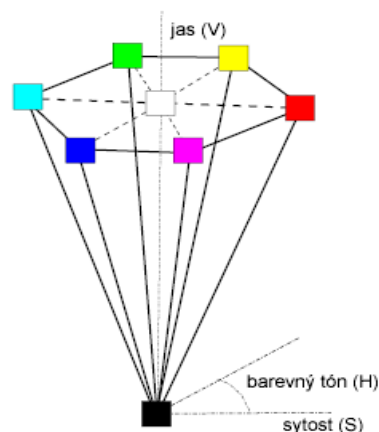
Převod mezi RGB a CMY prostory je možný na základě následující rovnice [19]:

$$\begin{pmatrix} C \\ M \\ Y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \quad (1.4)$$

kde R, G, B jsou kanály barevného prostoru RGB a C, M, Y jsou kanály prostoru CMY. [19], [21], [27]

1.1.3 HSV prostor

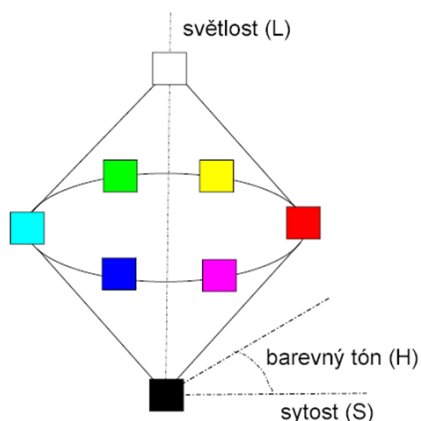
HSV prostor (někdy též uveden jako HSB - B -brightness) je pro intuitivní zadávání barev vhodnější než předchozí prostory. Prostor je znázorněn pravidelným šestistěnným jehlanem. Jednotlivé složky prostoru udávají následující parametry: H (hue) - barevný tón je určen převládající spektrální barvou, je zadán jako úhel s hodnotami od 0° do 360° . S (saturation) - sytost, určena jako relativní vzdálenost osy jehlanu (hodnoty v rozmezí $<0,1>$), dominantní barvy leží na plášti jehlanu (hodnota S je rovna jedné). V (value) - jasová hodnota, nejnižší jas je ve vrcholu jehlanu a směrem k podstavě stoupá, což je znázorněno na obrázku 1-3. Střed podstavy jehlanu představuje bílou barvu (maximální jas, maximální sytost). [19], [21], [27]



Obrázek 1-3 Zobrazení kombinace tří parametrů HSV prostoru (barevný tón, sytost a jas) a barev vycházejících z těchto parametrů [19]

1.1.4 HSL prostor

Je možné se setkat také s názvy HSI či HSD, tyto názvy označují stejný prostor jako HSL, třetí kanál je nazván odlišně: *L* (lightness), *I* (intensity), nebo *D* (darkness). Tento prostor je znázorněn pomocí dvojice kuželů spojených podstavami (viz Obrázek 1-4). Prostor je popsán pomocí tří parametrů: *H* (hue) - barevný tón nebo odstín, opět zadán jako úhel. *L* (lightness) - světlost, nejvyšší je u vrcholu spodního kuželu, naopak nejvyšší u vrcholu horního kuželu. *S* (saturation) - sytost, opět zadána jako relativní vzdálenost od osy kuželů. Tento model nejvíce odpovídá skutečnosti, že nejvíce barev člověk vnímá při průměrné světlosti, což na obrázku odpovídá oblasti spojených podstav dvou kuželů. Obrázek 1-4 znázorňuje zmíněné kužely a několik hlavních barev rozmístěných v objemu kuželů. [19], [21], [27]



Obrázek 1-4 Zobrazení kombinace tří parametrů HSL prostoru (barevný tón, sytost a světlost) a barev vycházejících z těchto parametrů [19]

1.1.5 YCbCr prostor

Formát YCbCr má opět tři kanály, první kanál Y prezentuje světelnost podobně jako u HSL na základě intenzity všech tří RGB komponent. Druhý a třetí kanál Cb a Cr reprezentují barvy, Cb komponenta je založena na základě rozdílu modré barvy (B z prostoru RGB) a světelnosti (Y), Cr komponenta je založena na rozdílu červené barvy (R z prostoru RGB) a světelnosti (Y). Zbývající informace pro zelenou barvu může být dopočítána z předchozích dvou komponent. Lidské oko je citlivější na změny ve světelnosti (Y komponenta) než na změny barev.

Převod obrazu z RGB prostoru do YCbCr prostoru je možný podle následující rovnice [54]:

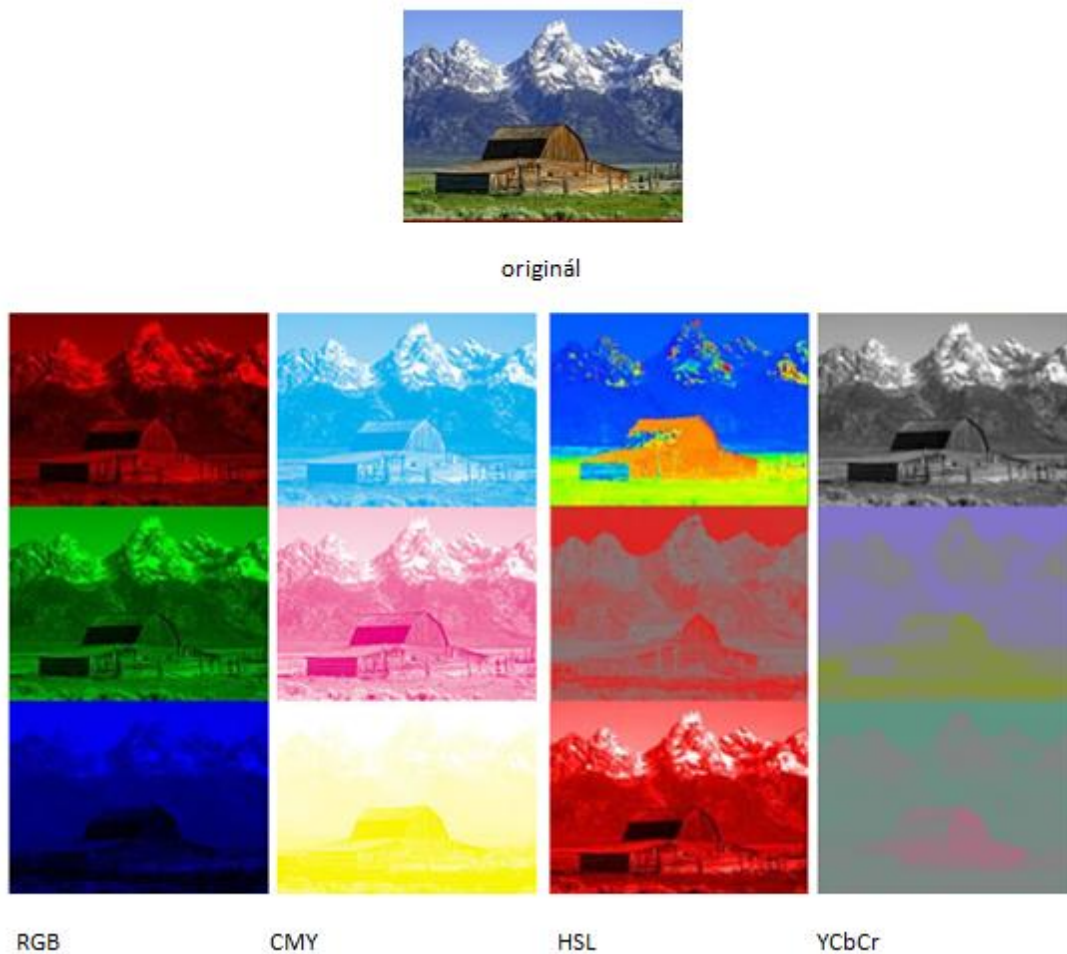
$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & -0.1689 & 0.4998 \\ 0.587 & -0.3317 & -0.4185 \\ 0.114 & 0.5006 & -0.0813 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \quad (1.5)$$

kde Y , Cb , Cr jsou kanály prostoru YCbCr a R , B , G jsou kanály prostoru RGB.

Tento formát je vhodný pro ukládání a přenos obrazových dat včetně videa. Využívají ho formáty JPEG (pro obrazy) a MPEG (pro video) při kompresi dat. Při přenosu jsou z důvodu citlivosti lidského oka informace o světelnosti komprimovány méně než informace o barvě.

Pro analogový signál lze použít prostor YPbPr odpovídající prostoru YCbCr pro diskrétní signál. Zmíněný prostor byl využíván u analogových televizních signálů. [19], [21], [50]

Uložení informace o barvě pixelů obrazu do několika kanálů různých barevných prostorů je znázorněno na obrázku 1-5. Je patrné, že v barevném prostoru RGB je vliv osvětlení ve všech kanálech, zatímco například u YCbCr prostoru osvětlení ovlivňuje pouze jeden kanál.



Obrázek 1-5 Zobrazení kanálů různých barevných prostorů [9]

1.1.6 YIQ

Y udává světelnost, I a Q jsou barevné složky. I udává barevný rozsah od oranžové barvy po azurovou, to je výhodou při detekce barvy obličeje u mongoloidní rasy. Q udává barevnou škálu mezi purpurovou a zelenou. Lidské oko je citlivější na změny v kanálu I než v kanálu Q , díky tomu může být kanál Q přenášen s menší přesností než kanál I . Převodem RGB prostoru na YIQ prostor dojde k potlačení pozadí ostatních barev a díky tomu je možné detekovat i obličeje relativně malé vůči celkové velikosti obrazu. Tento barevný prostor se využívá pro televizní signál ve Spojených státech. Televizi zobrazující pouze šedotónový obraz postačí pouze Y kanál.

Převod z RGB se realizuje podle následující rovnice [55]:

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.30 & 0.59 & 0.11 \\ 0.60 & -0.28 & -0.32 \\ 0.21 & -0.52 & 0.31 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \quad (1.6)$$

kde Y , I a Q představují kanály prostoru YIQ (světlost a dvě barevné složky), R , G , B

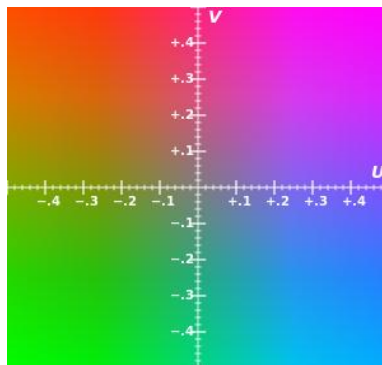
jsou tři barevné složky prostoru RGB (červená, zelená, modrá). [19], [21], [27], [55]

1.1.7 YUV

YUV prostor se v Evropě využívá pro televizní signál, oproti YIQ je barevný prostor pootočen o 33° . Převod z RGB prostoru je možný podle následující rovnice [55]:

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & -0.436 \\ 0.615 & -0.515 & 0.100 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1.7)$$

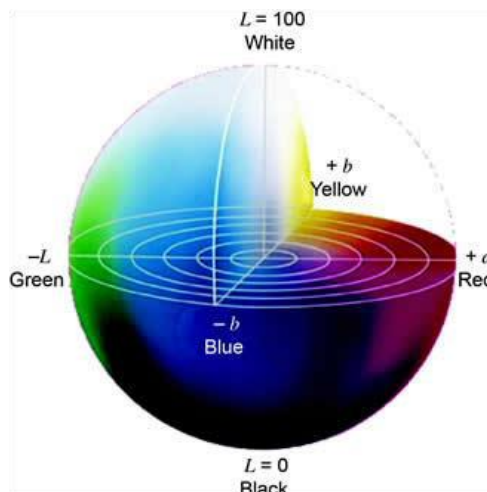
Obrázek 1-6 znázorňuje rozložení barevných odstínů v závislosti na hodnotách kanálů U a V . Z obrázku je zřejmé, že je barevný prostor vůči prostoru YIQ pootočen, u YUV udává kanál U barevný rozsah od zeleno oranžové barvy po modrou, kanál V barevný rozsah od zelené barvy po červenou. [19], [21], [27], [55]



Obrázek 1-6 Znáznornění kanálů U a V prostoru YUV udávajících barvu pixelu [55]

1.1.8 CIE L^*a^*b

CIE L^*a^*b je název prostoru, kde L značí osu světelnosti, rozpětí hodnot je od 0 do 100, přičemž nula odpovídá tmavému odstínu (černé barvě) a hodnota 100 odpovídá maximální světelnosti (bílé barvě). Na svislou osu L jsou kolmé osy a a b , osa a udává rozpětí os zelené barvy k červené, osa b udává rozpětí barev od modré po žlutou. Prostor L^*a^*b je znázorněn na obrázku 1-7. [19], [21], [27], [55]



Obrázek 1-7 Zobrazení prostoru CIE L^*a^*b [19]

1.1.9 Další barevné prostory

CIE-xyz - je numerický model senzitivity barvy vytvořený podle vnímání barev u lidí s normálním barevným viděním. Tento model reprezentuje průměrné barevné spektrum, které je pro člověka viditelné. Sítnice lidského oka obsahuje tři typy receptorů pro každou z barev: červená, zelená a modrá. Složky CIE-xyz prostoru odpovídají uvedeným barvám.

CIE L^*u^*v - je úprava prostoru CIE-xyz, tento prostor se navíc snaží dosáhnout percepční uniformity. Využívá se při práci s barevnými světelnými zdroji.

TSL - je barvený prostor, který definuje barvu pomocí odstínu, saturace a světlosti. Jelikož převod z RGB prostoru do TSL prostoru využívá normalizovaných kanálů r a g RGB prostoru, je TSL vhodný pro segmentaci barvy pokožky, neboť jeho složky T a S nejsou ovlivněny osvětlením.

YES - Y složka reprezentuje svítivost, je počítána jako vážená suma kanálů RGB prostoru. složky E a S udávají barvené faktory, E odpovídá rozdílu červené a zelené složky RGB prostoru, S je rovno rozdílu žluté a modré barvy.

[6], [9], [19], [21], [27], [53], [55]

1.2 Barevné formáty

V této kapitole jsou uvedeny používané barevné formáty, které slouží pro uložení obrazových dat. Jelikož tyto formáty zpravidla využívají kompresi dat pro snížení velikosti dat/souborů, jsou používané kompresní metody v této kapitole uvedeny také.

Ještě než budou jmenovány konkrétní kompresní metody a barevné formáty, je na místě uvést typy obrazů, se kterými se lze setkat. Obrazy lze rozdělit podle počtu bitů, které připadají na jeden pixel obrazu. Tento parametr se nazývá barevná hloubka. Mezi jednobitové obrazy patří černobílý obraz, kde pixel může mít pouze dvě barvy: černou nebo bílou. Osmibitovou barevnou hloubku mají šedotónové obrazy. Dvaceti čtyř bitová hloubka je u obrazů, jejichž pixely mají tři různé kanály. Tyto obrazy jsou barevné, příkladem mohou být RGB nebo YUV barevné prostory. Obrazy se čtyřmi kanály (RGBA, CMYK) mají třiceti dvou bitovou hloubku. Lze se setkat i s větší bitovou hloubkou (48 nebo 96 bitů), ale nejběžnější obrazy mívají bitovou hloubku 32 nebo 24 bitů. [19], [29], [31], [51], [54]

1.2.1 Kompresní metody

Pro uložení obrazů v určitém formátu se využívají kompresní metody, které sníží objem dat a tím sníží i nároky na paměť v místě uložení nebo v případě přenosu sníží nároky na rychlost přenosu dat. Kompresní metody lze rozdělit do dvou skupin podle zachování informace při kompresi dat. Bezeztrátová komprese při změně dat informace uchovává, naopak u ztrátové komprese dochází k ztrátě určité informace. V této kapitole budou uvedeny nejčastější metody komprese obrazů.

Run Length Encoding (RLE) je metoda kódování informace vhodná pro obrazy s většími plochami konstantních hodnot, nedochází ke ztrátě informací. V místě výskytu opakovaných hodnot na určitém řádku je jako první informace uložena hodnota pixelů, a jako druhá počet pixelů, které mají tuto hodnotu. Pokud jsou na řádku více než dva pixely za sebou shodné, bude komprese účinná a dojde k poklesu objemu dat. Pokud by obraz podmínky větších ploch nesplňoval, mohlo by dojít k negativní kompresi neboli k nárůstu dat. Této kompresní metody využívá formát PCX.

Huffmanovo kódování (CCITT) je opět bezeztrátová komprese, používána je například u formátu TIFF. Nejčastěji používané hodnoty jsou kódovány kódem s malou velikostí (1b), méně používané hodnoty jsou kódovány kódem o větší velikosti.

Slovníkové kódování (LZW) je dalším příkladem bezeztrátové komprese, využíváno je u formátů GIF, PNG, ZIP nebo ARJ. Využívá slovník se slovy o různých délkách. Delší slova vznikají kombinací kratších slov.

Diskrétní kosinová transformace (DCT) je převod jasových hodnot pixelů do

frekvenční oblasti, využívá ji formát JPEG. DCT není samo o sobě kompresní metodou, protože nezpůsobuje zmenšení objemu dat, ale po převodu dat do frekvenční oblasti a menších úpravách vysvětlených u formátu JPEG se ve frekvenční oblasti hodnoty velmi často opakují. Toho využívá Huffmanovo kódování, které u opakujících se hodnot v datech způsobuje zmenšení objemu dat.

[19], [25], [28], [29]

1.2.2 Používané barevné formáty

BMP (BitMaP) je formát vytvořený pro ukládání rastrové grafiky. BMP soubory jsou uloženy ve formátu nezávislém na zařízení, tedy způsobem, který není závislý na uspořádání pixelů nebo na interních metodách práce s barvou. BMP umožňuje ukládání dat ve čtyřech různých formátech: dvoubarevné obrázky (1 bit na pixel), mohou být vybrány libovolné dvě barvy, nemusí se jednat pouze o černobílý obraz. Dále šestnáctibarevné obrazy (4 bity na pixel), které využívají barevnou paletu o délce 64 bitů. Obrazy s barevnou hloubkou 256 (8 bitů na pixel), kdy má barevná paleta délku 1024 bitů. A nebo tak zvané Tru Color obrazy (24 bitů na pixel), které umožňují uložit 16 milionů barev, každý pixel je přímo reprezentován svou barvou a není třeba používat barevnou paletu.

GIF (Graphic Interchange Format) využívá bezztrátovou LZW kompresní metodu a tím velmi zmenší objem dat. Tohoto formátu se využívá pro přenos dat po síti, poskytuje možnost prokládání řádků. Pro rozpoznání obrazu není nutně potřeba celého objemu dat (stačí polovina nebo čtvrtina objemu dat). Lze také uložit více obrazů do jednoho souboru (zobrazení jako animovaný GIF). Tato skutečnost ale přináší jednu nevýhodu, formát je schopen uložit maximálně 256 barevných odstínů.

PNG (Portable Network Graphics) je taktéž primárně určen k přenosu dat po síti, může sloužit ale i pro archivaci dat. Využívá bezztrátovou kompresi algoritmu LZ77. Pixely jsou rozděleny do pěti skupin. Pixely první skupiny nejsou nijak upravovány, je uložena přímo jejich hodnota. Po pixely druhé skupiny (Sub) je ukládán rozdíl tohoto pixelu a pixelu vlevo od něj. Ve třetí skupině (Up) je ukládán rozdíl daného pixelu a pixelu na stejné pozici o řádek výše. Čtvrtá skupina (Average) ukládá průměr z daného pixelu a sousedních pixelů vlevo a nahoře. Poslední skupina (Paeth) ukládá hodnotu získanou z daných pixelů a tří sousedních (vlevo, nahoře a vpravo nahoře) metodou, kterou navrhl A. W. Paeth. Formát ukládá barevnou hloubku 24 bitů.

TIFF (Tagged Image File Format) je vhodný formát pro grafiku o velkém objemu dat. Podporuje 32 bitovou hloubku a bezztrátovou kompresi LZW. Je schopen uložit více obrazů do jednoho souboru.

JPEG - Jelikož obraz je schopen tolerovat určitou míru ztráty dat, může být

efektivně využita ztrátová komprese. Nejvíce používanou ztrátovou kompresí je technika JPEG komprese. (Zkratka JPEG vyjadřuje Joint Photographic Experts Group.) JPEG se nejvíce hodí pro fotografie a podobné obrazy, které jsou charakterizovány jemnými detaily a kontinuální změnou tónu barev. Obraz je nejprve převeden do YCbCr prostoru, transformace je realizována podle rovnice uvedené v kapitole 1.1.5.

Dále je každá složka obrazu v YCbCr prostoru je podvzorkována a zvlášť zpracovávána. Každá složka je převedena na frekvenční reprezentaci pomocí diskrétní kosinové transformace. Vstupem do výpočtu diskrétní kosinové transformace je matice pixelů, výstupem je matice koeficientů, reprezentujících amplitudy frekvenčních komponent obrazu. Komprese i dekomprese touto metodou je poměrně výpočetně náročná. Komprese se neprovede na celém obraze, ale obraz je nejprve rozdělen na čtverce o straně 8 pixelů a každý z těchto čtverců je převeden na frekvenční reprezentaci zvlášť. Převod obrazu do frekvenční oblasti nepřináší žádnou kompresi objemu dat, ale transformuje data do podoby, která může být komprimována způsobem, který minimalizuje možnost poznat ztrátu dat. Informace o vysokých frekvencích mohou být zanedbány, protože vysoké frekvence příliš nepřispívají ke vnímané kvalitě obrazu. Budou ponechány například jen hodnoty vysokých frekvencí dělitelné číslem 4 beze zbytku, a to pouze hodnoty do určité meze. Tím se výrazně sníží počet hodnot pro vysoké frekvence. Čísla kvantizačních úrovní bývají uvedena v kvantizační matici, výsledky diskrétní kosinové transformace jsou vyděleny kvantizační maticí a následně zaokrouhleny. Kvantizační přístup redukuje velikost uložených dat. Jelikož se ve výsledné matici čísla (především nuly) často opakují, využívá se Huffmanova kódování, které snižuje objem dat. Aby komprese byla co nejúčinnější, jsou koeficienty výsledné matice zpracovány takzvanou zig-zag metodou, která maximalizuje počet opakování nulových hodnot za sebou v jedné linii.

Před zobrazením obrazu je postup opačný, nejprve je provedena dekomprese bezetrátových metod, které byly použity při kompresi. Následuje násobení bloků 8x8 pixelů kvantizační maticí a inverzní kosinová transformace. Bloky odpovídající jednotlivým složkám obrazu (*Y*, *Cb*, *Cr*) jsou sloučeny do matic. Jelikož při kompresi došlo ke snížení objemu dat, je nyní nutné chybějící data dopočítat, aby matice měly stejný rozměr jako původní obraz. Poslední operací je převod obrazu z YCbCr prostoru do RGB.

JPEG formát je standard a jeho kodéry a dekodéry jsou implementovány ve všech běžně používaných zařízeních jako jsou počítače, telefony, tablety a další.

[15], [25], [28], [29], [32], [40], [54]

2 DETEKCE OBLIČEJE V OBRAZE

Před tím než bude možné vyhledávat jednotlivé objekty či rysy obličeje, je třeba stanovit oblast obličeje, kde se budou tyto objekty hledat. Sníží se tím nejenom výpočetní náročnost algoritmu, ale i pravděpodobnost chybné detekce, například detekce objektů v pozadí mimo obličej. V této kapitole budou uvedeny různé algoritmy pro detekci tváře. Detekce tváře je využívána například pro rozpoznávání obličejů za účelem identifikaci osob, zpracování obrazů na základě jejich obsahu, videokonference nebo počítačové vidění. Lidská tvář může být velmi různorodá co do tvaru, barvy pleti a rozmístění objektů obličeje, což znesnadňuje automatickou detekci tváře například pro počítačové vidění.

Metody detekce obličeje lze dělit podle různých aspektů. Podrobněji rozepsáno bude rozdělení na základě přístupu, jakým uvedené metody využívají informace z obrazu.

V tomto rozdělení jsou první skupinou **příznakové metody** (**Feature-based methods**). Metody spadající do této kategorie využívají znalosti o tváři přímo a používají klasické metody detekce založené na primárních parametrech jednotlivých pixelů nebo menšího okolí daného pixelu. Hledání určitých příznaků v obraze mohou předcházet jednoduché operace, díky kterým je získána jiná reprezentace obrazu, například hranová reprezentace určená z malého okolí každého pixelu. Pro detekci tváře se často využívá měření vzdáleností mezi příznaky, úhlů mezi nimi nebo plocha příznaků v obraze. Metody této kategorie mohou využívat také například znalost tvaru obličeje nebo znalost barvy pleti. Je logické, že tyto metody byly nejvíce publikovány v sedmdesátých letech, když bylo počítačové vidění, detekce určitých objektů v obraze a podobné úlohy na počátku svého rozvoje. Přednostmi metod této skupiny je relativní jednoduchost algoritmů jednotlivých metod a s tím spojená vyšší rychlost algoritmu v porovnání s druhou skupinou metod. Příznakové metody fungují bez jakýchkoli znalostí nebo informací, které musí být vloženy před samotnou detekcí obličeje.

Druhou skupinou jsou **metody založené na obrazové informaci** (**Image based methods**). Tyto metody přistupují k detekci tváře jako k obecnému problému a nahlíží na obraz jako na jeden celek. K detekci obličeje v obraze využívají množinu obrazů se vzorovými obličejí a proces učení algoritmu. Vstupní obraz je pak přímo klasifikován na části odpovídající obličejí a na části, kde se obličej nevyskytuje. Algoritmus pracuje buď s celým vstupním obrazem, nebo s jeho výřezy, a bere pouze jasovou informaci obrazu, která byla vstupem algoritmu. Na rozdíl od první skupiny metod zde nejsou

získávány žádné další příznaky obrazu. Jelikož jsou metody této skupiny poměrně nové, byly jednotlivé metody publikovány později než metody první skupiny. Metody této skupiny mohou dávat přesnější výsledky než příznakové metody. Ovšem tyto metody jsou založeny na výrazně složitějších algoritmech a často vyžadují obecné apriorní znalosti vložené před detekcí v konkrétních obrazech. Další nevýhodou může být časově náročné učení algoritmů.

Vedle rozdělení na základě přístupu k informacím obrazu lze metody také rozdělit podle velikosti oblasti, ze které budou brány informace, podle nichž bude detekce obličeje provedena. Při detekci každého jednotlivého pixelu je na základě rozhodovacího pravidla dané metody rozhodnuto, zda tento pixel odpovídá hledané oblasti či ne. Pokud bude v potaz bráno určité okolí každého pixelu, v každém okolí bude rozhodnuto, zda se v něm nalézá hledaný objekt. Velikost okolí může být během procesu detekce postupně zvětšována (nebo zmenšována) v závislosti na velikosti obličeje v obraze. Výstupem bude obdélník, který v obraze vyznačí oblast, v níž se hledaný objekt nachází. Obdobnou metodou je využití podvzorkování obrazu, kdy je obraz vždy převzorkován na poloviční velikost. Díky tomu lze také zachytit v obraze obličeje různých velikostí.

Jednotlivé metody budou představeny v následujících podkapitolách. Jsou řazeny podle přístupu, jakým využívají informace z obrazu. [6], [8], [12], [16], [17], [21], [22], [26], [34], [56]

2.1 Příznakové metody (Feature-based Approach)

Příznakové metody lze rozdělit do tří skupin. Při hledání obličeje v obraze jsou jako nejjednodušší postup aplikovány jednoduché metody zpracování obrazu, které na výstupu dávají některou ze základních charakteristik obrazu (Low-level Analysis). Metody využívají vlastností jednotlivých pixelů (například stupeň šedi či barva pixelu) nebo vlastnosti lokálního okolí pixelu, a výstupem je silueta objektu. Výsledky mohou být nejednoznačné kvůli skutečnosti, že informace pochází pouze z hodnot jednoho pixelu nebo malého okolí pixelu a chybí komplexní pohled na vlastnosti obrazu. U metody hledání rysů obličeje (Feature Analysis) je tvář brána z více globálního pohledu s ohledem na její tvar. Díky tomu je nejednoznačnost výsledků v porovnání s analýzou základních charakteristik obrazu výrazně nižší a lze lépe určit pozice obličeje v obraze. Třetí skupina využívá aktivních kontur (Active Shape Model), vyhledává komplexní prvky obličeje, jako jsou rty nebo zornice. [21], [26], [56]

2.1.1 Metody využívající základní charakteristiky pixelu nebo jeho malého okolí (Low-level Analysis)

Každá z metod této podskupiny využívá některou ze základních charakteristik pixelu, případně jeho malého okolí. Metody v této kapitole jsou řazeny podle těchto charakteristik.

Hrany - Za jeden z nejzákladnějších příznaků obrazů se dá považovat hranová reprezentace. Pomocí sledování hrany (s omezením zakřivení kvůli omezení sledování falešných hran) lze nalézt objekty v obraze, v tomto případě především objekty obličejů. Před samotným sledováním hran je nutno provést úpravu hran (ztenčení, propojení a pod.). Na hrany uvnitř obličejů je aplikována příznaková analýza s využitím znalosti tvaru a pozice hlavy. Pomocí této metody lze v obraze rozpoznat například brýle na obličejích. Pro hranovou detekci se používají různé operátory, mezi často užívané operátory patří Sobelův operátor nebo LoG (Laplacian of Gaussian).

Jasová informace šedotónového obrazu – Jasovou informaci lze také využít pro detekci obličejů v obraze. Některé objekty obličejů jako například obočí, rty nebo zornice mají nižší jasové hodnoty než okolní kůže. Takto mohou být rozpoznány jednotlivé části obličejů. V oblasti obličejů jsou prahováním vyhledávána lokální minima jasu, která odpovídají zmíněným objektům obličejů. Lokální maxima (světlý pixel obklopen osmi tmavými pixely) v oblasti obličejů odpovídají světlým oblastem obličejů, například špičce nosu. Detekovaná místa jsou srovnána s šablonou ideálního obličejů. Před samotnou detekcí je obraz na vstupu nejprve upraven např. ekvalizací histogramu nebo úpravou kontrastu pro lepší detekci lokálních jasových minim a maxim.

Jiný přístup využívající šedotónovou informaci je založen na skutečnosti, že pokud bude sníženo rozlišení obrazu, zaniknou mikroskopické příznaky a oblast obličejů se stane jednotnou. Uniformní oblasti jsou hledány pomocí skupiny pravidel určující jednotnost oblasti, ověření uniformní oblasti je možné pomocí existence typických rysů obličejů v těchto oblastech získaných jako lokální minima z obrazu s vysokým rozlišením (stejný princip jako je uveden výše).

Jasová informace barevného obrazu - V porovnání s šedotónovým obrazem přináší barevný obraz více informací, neboť určitá jasová hodnota šedotónového obrazu odpovídá více různým barevným odstínům barevného obrazu. Pokud by barva pozadí byla předem známá, je detekce hlavy, případně těla, před pozadím jednoduchá operace. Pokud bude obličej hledán podle barvy kůže, je důležité vědět, že variace barvy lidské pleti se pohybuje v určitém prostorovém shluku a rozdíly barvy kůže u jednotlivců (včetně různých ras) jsou poměrně malé. Pro barvu kůže existují různé podmínky v závislosti na zvoleném barevném prostoru. Je vhodné použít prostory, kde osvětlení objektu ovlivňuje pouze jednu složku, a další dvě složky jsou na osvětlení nezávislé, pro

tyto dvě složky budou určeny podmínky, za kterých daný pixel odpovídá barvě kůže. V případě použití barevného prostoru, ve kterém osvětlení ovlivňuje všechny složky, je vhodné před detekcí vliv osvětlení odstranit. Barevné prostory jsou uvedeny v kapitole č.1.1.

Obecně může být segmentace na základě barvy kůže provedena pomocí stanovení určitých hranic, které rozliší pixely odpovídající barvě kůže od ostatních pixelů. Barva kůže může být zobrazena v grafu nebo histogramu, odkud lze odhadnout zmíněné hranice. Tuto metodu lze kromě nalezení obličejů využít také pro rozeznávání příkazů zadaných pohybem ruky.

Tato metoda není příliš robustní, problémy mohou být způsobeny osvětlením scény nebo různorodostí barvy lidské kůže. Nicméně implementace této metody je jednoduchá a algoritmus je poměrně rychlý. Za určitých podmínek nemusí být nevýhody této metody zásadní, nebo se dá těmto nevýhodám předejít. Jelikož byla tato metoda vybrána pro řešení detekce tváře a implementována v prostředí Matlab, bude podrobněji popsána v kapitole č.4.

Pomocí pohybu – Informaci o pohybu dostupnou z videa lze využít k nalezení pohybujících se objektů. Ve videu lze odlišit pohybující se popředí (osobu) vůči statickému pozadí. Pohybující se objekty jsou extrahovány odečtením dvou obrazů s rozdílným časovým parametrem (zbude pouze to, co je na obrázcích odlišné, tedy objekty, které v časovém intervalu mezi dvěma snímky změnily polohu). Tento princip je jednoduchý a efektivní, není ovlivněn charakterem pozadí. Pokud se ovšem ve videu vyskytne více pohybujících se objektů, z nichž některé neodpovídají hledaným osobám nebo obličejům, tento algoritmus selhává. V takovém případě by bylo třeba tuto metodu zkombinovat s jinou metodou, která odliší falešné pohybující se objekty od objektů, které má metoda skutečně detekovat. Při detekci očí v obličejích může být využita skutečnost, že oči při natáčení videa mrkají (blinking detection).

Jiný způsob, jak nelézt pohybující se objekty, je metoda s využitím odhadu kontur pohybujících se objektů. Výsledky v porovnání s předchozí metodou využívající video jsou věrohodnější, zřejmě je to především, pokud je pohyb objektů ve videu nepatrný. [21], [24], [34], [53], [60]

2.1.2 Metody hledání rysů obličeje (Feature Analysis)

Analýza využívající základní charakteristiky jednotlivých pixelů nebo malého okolí pixelu dává nejednoznačné výsledky. Příkladem mohou být u detekce na základě barvy kůže chybně stanovené regiony kůže v pozadí, kde se barva podobá barvě kůže. Metody hledání rysů tyto problémy nejednoznačnosti odstraní. Je využívána znalost tvaru obličeje, detekce obličejových částí je porovnána s geometrií tváře a chybně

detekované oblasti jsou rozpoznány a označeny jako mimobličejové části.

V následujících odstavcích budou popsány přístupy využívající znalost geometrie tváře. První přístup zahrnuje různé způsoby hledání rysů obličeje založené na relativní pozici rysů vůči celému obličeji. Důvěryhodnost detekce jednotlivých rysů je zvyšována nalezením dalších rysů v okolí. Druhý přístup využívá flexibilního uspořádání rysů v obličeji s využitím různých modelů obličeje.

První přístup – hledání rysů - Nejprve se určí významné rysy obličeje. Při hledání dalších méně významných rysů jsou využívány informace o pozici významných rysů, což usnadňuje detekci ostatních rysů. Správnost detekce je ověřena na základě geometrického rozložení nalezeného rysu vůči významným rysům obličeje. Významným rysem mohou být například oči, využívá se jejich symetrie a víceméně neměnné velikosti vůči obličeji.

Příkladem hledání rysů může být algoritmus, který zveřejnil De Silva a kolektiv v [33], kde je nejprve určena oblast, která bude pokládána za vrcholek hlavy, a postupným prohledáváním směrem dolů od vrcholku hlavy se hledá rovina očí, která se vyznačuje vyšším počtem hran (v hranové reprezentaci vstupního obrazu) než okolí. Vzdálenost roviny očí od vrcholku hlavy bude brána jako referenční a bude využita pro nastavení flexibilní obličejové masky. Tento algoritmus lze použít pro snímky s čelním postavením obličeje nebo snímky s natočením tváře o méně než 30°. Nevýhodou algoritmu je, že nedokáže správně detekovat jednotlivé rysy obličeje, pokud se na obličeji vyskytují brýle, nebo pokud je čelo zakryto vlasy.

Druhý přístup - konstelační analýza (constellation analysis) - Zatímco některé metody zmíněné výše jsou závislé na heuristických informacích snímků pořízených za určitých podmínek a nejsou proto dostatečně robustní k obecnému hledání obličejů v jakékoliv scéně, druhý typ přístupů seskupuje příznaky obličeje do kombinace pomocí robustnějších metod, jako například statistická analýza.

Pro výskyt jednoho i více obličejů v obraze může být použit pravděpodobnostní model tváře. Jednotlivé rysy tváře jsou získány pomocí hranové reprezentace. Rysy jsou zařazeny do různých skupin rysů, což umožňuje větší flexibilitu detekce a redukuje chybné detekce rysů. Algoritmus je schopen správně detekovat tváře v obraze i bez některých chybějících rysů. Do jisté míry je schopen se vypořádat i s problémy vzniklými posunutím, natočením nebo změnou měřítka obličeje. Úspěšnost tohoto postupu je asi 92%, největší počet chyb detekce tváře způsobuje větší natočení hlavy ve snímcích. Algoritmus je schopen správně detekovat i za přítomnosti brýlí na obličeji, několik chybějících rysů také nezpůsobí chyby detekce a ani malé změny v natočení nebo měřítku neznemožní správnou detekci rysů obličeje.

Jiný způsob může být založen na využití Houghovy transformace pro hledání tvaru

obličej. Obraz je nejprve pomocí gradientního operátoru převeden na čelní zobrazení. Pomocí Houghovy transformace jsou hledány objekty elipsovitého tvaru, které by mohly odpovídat lidským obličejům. Tyto objekty jsou následně poskytnuty vzorové šabloně, která potvrdí nebo vyvrátí skutečnost, že dané objekty odpovídají lidské tváři.

[21], [33], [34]

2.1.3 Metoda aktivních kontur

Aktivní modely zobrazují skutečný fyzikální vzhled rysů. V okolí každého pixelu jsou hledány lokální rysy. V těsném okolí určitého rysu bude aktivní tvarový model interagovat s lokálními příznaky obličeje (např. hrany a jas) a bude se postupně deformovat, aby se přiblížil tvaru daného rysu. Existují tři typy aktivních tvarových modelů pro vyhledávání rysů v obličeji. První typ využívá obecnou tvarovou konturu, které se říká snake (had). Druhý typ využívá deformovatelnou šablonu a třetí typ pracuje s konturou sestavenou z určitého počtu bodů rozmístěných v prostoru (PDM).

Snakes - pro vyhledání hranice hlavy, na začátku je had inicializován v okolí hranice hlavy, pak se přibližuje hranám v jeho okolí a dostává tvar hranice hlavy. Vývoje hada je dosaženo minimalizací funkce energie hada. Nedostatek této metody spočívá v tom, že často část hada přilne ke špatnému rysu obličeje. Had také není schopen přilnout ke tvaru nekonvexních objektů, což je způsobeno minimalizací zakřivení. Výhodou hada ale je, že na rozdíl od jiných zmíněných metod dokáže detekovat brýle na obličeji.

Deformovatelná šablona - Stejně jako je had schopný se deformovat podle tvaru objektů v jeho blízkosti, lze použít i masku, která se bude deformovat. A sice na základě nejstrmějšího sestupu kombinace několika lokálních charakteristik (a sice na základě energie, hran a jasu). Jako příklady lze uvést šablonu oka, nebo úst.

Point Distributed Models (PDM) - Kontura se sestává z určitého počtu bodů, kterým jsou přiřazeny prostorové souřadnice. Variace těchto bodů je napřed provedena na trénovací množině objektů (s různým tvarem, velikostí a s různým umístěním), přičemž jsou získány variace velikosti tváře a tvaru významných rysů (oči, rty a jiné). Pomocí analýzy hlavních komponent (PCA) jsou variace rysů v trénovací množině použity jako lineární flexibilní model. Model zahrnuje průměr všech rysů trénovací množiny a způsoby změn pozice každého bodu. Výhodou PDM je fakt, že metoda dává kompaktní parametrický popis. Může být použita i pro interpretaci výrazu tváře. [21], [47]

Aktivní vzhledový model (AAM Active Appearance Model) - Tento model je zobecněním metody aktivních kontur, využívá více informací z celé oblasti zájmu

obrazu, na rozdíl od metody aktivních kontur, která využívá pouze informace z nejbližšího okolí hran nebo vybraných bodů. Model obsahuje informace o tvaru tváře, která je uložena pomocí bodů spojených do 2D sítě trojúhelníků, a o její textuře, vyjádřené jasnem jednotlivých pixelů uvnitř trojúhelníkové sítě. V trénovací množině obrazů jsou snímky s ručně vyznačeným výskytem tváře s požadovaným počtem klíčových bodů. Snímky jsou upraveny tak, aby na všech byly tváře stejné velikosti a se stejným natočením. Následně je vypočítán střední tvar modelu a pomocí PCA je vypočtena transformační matice a vektor parametrů, který definuje povolenou deformaci modelu. Obdobně se vypočítá i střední textura modelu, opět se vychází z transformační matice a z vektoru parametrů pro povolené změny v textuře.

Se znalostí středního tvaru a střední textury je možné vytvořit nové tváře modelu, a sice pomocí afinní transformace, kdy je každému pixelu středového tvaru vypočtena nová pozice. Metoda umí generovat tváře s rozlišením pohlaví nebo věku, pokud jsou tyto rozdíly zahrnuty do trénovací množiny obrazů.

Při nasazení modelu na nový snímek je požadována minimalizace odchylky mezi novým snímkem a snímkem modelu (nejlepší je využití středního modelu), minimalizace respektuje povolenou deformaci tvaru a povolenou změnu textury.

[14], [21], [37], [47], [56], [63]

2.2 Metody založené na obrazové informaci (Image-based Approach)

Modely rysů obličeje narážejí na nepředvídatelnost podoby tváře a okolní podmínky, což velmi znesnadňuje nalezení obličeje. Většina zmíněných příznakových metod vyžaduje více či méně čelní postavení obličeje. Příznakové metody nejsou schopny spolehlivě nalézt větší počet tváří v různorodém pozadí.

Tento problém se snaží vyřešit metody založené na obrazové informaci, u nichž není vyžadována znalost tváře (specifických rysů a pod.). Tím se sníží potenciál chyb vznikajících kvůli neúplným či nesprávným informacím o tváři.

Základem metod založených na obrazové informaci je trénování algoritmu, kdy jsou obrazy trénovací množiny zařazeny do jedné ze dvou tříd: "s obličejem" nebo "bez obličeje". 2D matice intenzit vstupního obrazu je pak porovnána s těmito dvěma třídami a podle toho je rozhodnuto o výskytu obličeje ve vstupním obraze. Nejjednodušší algoritmy založené na obrazové informaci spočívají v porovnávání se vzory (templatematching), ale jejich výsledky nejsou vždy stoprocentně správné.

Většina metod patřících do této kategorie používá k detekci tváře postupné zpracování částí obrazu pod posouvajícím se oknem. V části, která se právě nalézá pod posuvným oknem, proběhne kompletní prohledávání možného výskytu tváře a je rozhodnuto, jestli se ve výřezu hledaná tvář nachází či nikoli. Různé algoritmy se odlišují velikostí okna, ve kterém dochází k prohledávání, mírou podvzorkování obrazu, překryvem oken či počtem iterací. Velikost okna se může během průběhu algoritmu postupně měnit. Výstupem metod založených na obrazové informaci je obdélník vyznačující oblast (včetně její velikosti), kde se hledaný objekt nachází.

Metody založené na obrazové informaci lze rozdělit do tří kategorií: lineární podprostorové metody, neuronové sítě a statistické metody. [14], [21], [23], [56]

2.2.1 Lineární podprostorové metody

V této skupině technik se lze často setkat s využitím PCA - analýzou hlavních komponent, LDA - lineární diskriminační analýzy, nebo s využitím faktorové analýzy (FA). V případě využití PCA jsou na soubor různých obrazů lidského obličeje nejprve nalezeny hlavní komponenty distribuce tváří, udány jsou ve vektorech vlastních hodnot (vlastní hodnoty kovariační matice distribuce). Každá z tváří souboru může být aproximována lineární kombinací největších vektorů vlastních hodnot s příslušnými vahami. Tato metoda může být použita i pro rozpoznávání obličejů.

Detektor rysů obličeje může využívat DFFS (distance-from-face-space, indikátor existence tváře pomocí pozorování globálního minima distanční mapy) generovaného z vlastních vektorů (vlastní vektory očí, nosů, úst) trénovací množiny obrazů. Tento detektor je schopný detekovat tváře pod různými úhly, v případě že trénovací množina obsahuje obrazy také pod různými úhly.

PCA je vhodný intuitivní způsob vytvoření podprostoru reprezentujícího objekt určité třídy. Nicméně PCA není nutně vždy optimální. Oblast obličeje může být lépe reprezentována rozdělením do podtříd. Takové metody využívají obvykle kombinace vícerozměrných Gaussianů. [12], [21], [56]

2.2.2 Neuronové sítě

Neuronové sítě jsou často používány k různým úlohám zahrnujícím rozeznávání objektů, kromě hledání lidské tváře mohou vyhledávat a klasifikovat jednotlivé rysy tváře. U detekce lidské tváře je využita jasová reprezentace a následuje korekce osvětlení a vyvážení histogramu. Neuronová síť posuzuje okno o určité velikosti (např. 20x20 pixelů). Pro detekci tváře může být neuronová síť navržena s jednou skrytou vrstvou a určitým počtem neuronů v ní, které se zabývají různě velkými podoblastmi

okna. Pro dosažení lepších výsledků lze použít více neuronových sítí a jejich výstupy zkombinovat například logickými operátory AND či OR, nebo použít výsledek prezentovaný největším počtem neuronových sítí. [12], [14], [21], [56], [62]

2.2.3 Statistické metody

Tyto metody popisují obrazy s charakteristickými obličejí pomocí statistiky, například pomocí histogramu nebo invariantního momentu. Algoritmy mohou pro extrakci rysů obličeje využít odčítání po sobě jdoucích výřezů, autokorelace nebo invariantní moment. Výhodou metody je to, že nalezení rysů obličeje není ovlivněno rotací, translací nebo změnou měřítka.

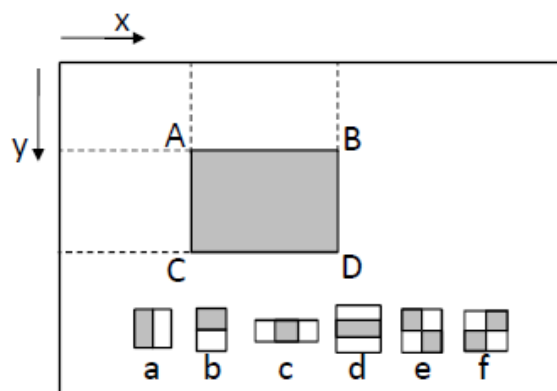
Mezi statistické metody patří metoda založená na Bayesově síti, což je pravděpodobnostní grafický model založen na Bayesově vzorci, systémy založené na informační teorii nebo metodě podpůrných vektorů SVM (Support Vector Machine). SVM patří mezi metody strojového učení a je založena na principu strukturální minimalizace rizik klasifikační metody. SVM s jádrovou funkcí tvaru polynomu druhého stupně je trénován s dekompozičním algoritmem, který zaručuje dosažení globálního optima. Trénovací data jsou zaříděna do kategorií podle prostorového rozmístění v nadrovině. [14], [21], [60], [62], [67]

2.2.4 Viola-Jones algoritmus

Viola-Jones detektor se používá pro detekci tváře, řadí se mezi metody založené na obrazové informaci a vyhledává předem určené objekty. Algoritmus využívá tři základních částí, které umožňují označit tento postup jako real time, neboli proces, který lze aplikovat v reálném čase a mohl by být použit i ve videu. Jsou to výhody integrálního obrazu, učení klasifikace pomocí metody Adaboost a kaskádová struktura klasifikace.

Kaskáda klasifikátorů na základě informativních Haarových příznaků rozhoduje o přítomnosti tváře v aktuální oblasti obrazu. Kaskádou klasifikátoru se rozumí určitý počet jednoduchých klasifikátorů, kterými postupně prochází aktuální oblast obrazu, dokud není u některého klasifikátoru oblast vyhodnocena jako oblast bez přítomnosti tváře a klasifikace je ukončena, nebo dokud oblast neprojde všemi klasifikátory. Jednotlivé klasifikátory jsou tvořeny rozhodovacími stromy alespoň se dvěma úrovněmi. Rozhodování v každém klasifikátoru se provádí jednou ze čtyř následujících pomocných metod: Discrete Adaboost, Real Adaboost, Gentle Adaboost nebo Logitboost. Informativní Haarovy příznaky jsou výstupy Haarových filtrů, jsou to čísla, která odpovídají rozdílu sum jasových hodnot pod světlou a pod tmavou částí Haarova

filtru. Příklady masek Haarových filtrů jsou uvedeny na obrázku 2-1 pod písmeny a až f.



Obrázek 2-1 Znázornění posuvu okna po obraze a příklady masek Haarových filtrů (a-f) [42s]

Rozhodování o přítomnosti hledaného objektu probíhá postupně na výřezech obrazu a každý výřez je označen jako hledaný objekt (v tomto případě se jedná o tvář), nebo jako pozadí. Výřezem se rozumí oblast, která se v dané chvíli nachází pod oknem o určité velikosti. Okno se v průběhu algoritmu postupně posouvá, až projde celou plochu obrazu. Posouvání okna po vstupním obraze je naznačeno na obrázku 2-1 šedým obdélníkem ABCD. Velikost okna a tím i velikost výřezu se může v průběhu algoritmu postupně zmenšovat nebo zvětšovat. Využití informativních Haarových příznaků výrazně urychluje tento algoritmus. Kromě toho se pro urychlení také využívá sumačního obrazu, čímž se sníží výpočetní náročnost a počet přístupů k paměti procesoru. Konkrétní Haarovy filtry, a tedy i Haarovy příznaky, které budou použity, jsou vybrány v procesu učení, které je zprostředkováno například metodou Adaboost. Učení probíhá na malých obrazech o velikosti například 19x19 pixelů. Musí být zastoupeny obě skupiny obrazů, to znamená skupina s výskytem tváře v obraze i skupina obrazů bez tváře.

Metoda Adaboost - Její název je zkratkou slov adaptive boosting. Pokud předchozí klasifikátory chybně rozpoznaly hledaný objekt, dochází u v následujícím kroku ke zlepšení klasifikátoru. Metoda je citlivá na odlehlá data a na šum. Po trénování klasifikátorů se provádí výpočet chybovosti klasifikátorů a nejméně chybový klasifikátor je zařazen do množiny základních klasifikátorů. Tento postup se opakuje pro různé nastavení vah vstupujících do výpočtu.

[4], [17], [41], [47], [56], [57], [63], [64]

2.2.5 Skrytý Markovský model

Skrytý Markovský model využívá Markovský proces, který obsahuje skryté neznámé parametry, a dokáže efektivně popsat statistický model náhodné informace. Tento model se dá využít k rozpoznání řeči, nebo rozpoznání obličejů nebo ke klasifikaci pohybu ve videu. Výhodou metod rozpoznávajících obličej je, že je odolná vůči změně rysů obličejů a vůči natočení hlavy. [21], [49]

Úspěšnost detekce metod založených na obrazové informaci se pohybuje mezi 73 a 95% v závislosti na zvolení konkrétní metody a jejích parametrů. [21] Vzhledem k tomu, že úspěšnost metod byla testována na velmi komplexní databázi obrazů, obsahující barevné i šedotónové obrazy, rozmanité úhly natočení hlav osob, komiksové obrazy nebo obrazy kreslené rukou, dají se dosažené výsledky považovat za velmi dobré. Na rozdíl od příznakových metod zde nejsou překážkou detekce ani brýle na obličej, menší části obličejů zakryté vlasy nebo výskyt více obličejů v jednom obraze.

Detekci obličejů v obraze mohou komplikovat proměnlivé parametry v obrazech, jako jsou například různé pozice a natočení tváře (pohled zepředu, z profilu, aj.), osvětlení scény, orientace obrazu, různé výrazy tváře. Pokud všechny vstupní obrazy spojují určité vlastnosti, například podobné osvětlení nebo čelní postavení hlavy, mohou být použity jednodušší algoritmy a detekce bude méně časově náročná. Také vhodně zvolená kombinace některých uvedených metod může dát lepší výsledky než jednotlivé metody. [12], [21], [59], [66]

2.3 Morfologické operace

Metody detekce obličejů bývají často doplněny o morfologické operace, které pomáhají pročistit detekované oblasti v binárním obraze tak, aby výsledné oblasti detekce odpovídaly oblastem obličejů. Cílem morfologických operací v tomto případě je odstranit malé detekované oblasti, které neodpovídají oblastem obličejů, a odlišit od sebe jednotlivé objekty v binárním obraze. Objekty by měly být jednolitě a měly by mít hladký obvod.

Morfologické operace obecně označují nelineární operace související s tvarem nebo

s morfologií příznaků zpracovávaného obrazu. Primárně jsou morfologické operace určeny pro binární obrazy, lze je však aplikovat i na obrazy šedotónové. Jedná se o zpracování obrazu pomocí masky nazývané strukturní element. Maska obsahuje pouze hodnoty jedna a nula (lze ji chápat jako malý binární obraz). Morfologické operace se řadí mezi lokální metody, obraz je zpracováván postupně po oblastech, které se aktuálně nachází pod maskou posunující se po obraze. Velikost masky je řádově menší než velikost obrazu.

Základními morfologickými operacemi jsou eroze a dilatace, otevření a uzavření jsou kombinacemi prvních dvou operací. [38], [42], [43], [45], [52]

2.3.1 Eroze

Eroze je základní morfologická operace, která zmenšuje velikost objektů obrazu, odstraňuje malé objekty či ostrůvky vycházející z větší oblasti. Výstupní obraz má hodnoty jedna na souřadnicích, kde jsou hodnoty části vstupního obrazu překryté maskou a hodnoty masky shodné. Na ostatních souřadnicích výstupního obrazu jsou nuly. Operace je popsána v rovnici 2.1

$$B \ominus S = \{b \mid (b + s) \in B, \forall s \in S\} , \quad (2.1)$$

kde B představuje vstupní binární obraz, S je strukturní element (maska), b a s značí jednotlivé pixely matice B případně S . [38], [42]

2.3.2 Dilatace

Dilatace představuje zvětšování objektů obrazu, využívá se k vyplnění malých děr uvnitř větších oblastí. Pokud se středový pixel masky shoduje se středovým pixelem překryté části vstupního binárního obrazu, získávají pixely výstupního obrazu na aktuální pozici masky hodnotu jedna. V opačném případě mají tyto pixely hodnotu nula. Definice dilatace je uvedena v rovnici 2.2

$$B \oplus S = \{(b + s) \mid b \in B, s \in S\} , \quad (2.2)$$

Dilatace je asociativní a komutativní operace, eroze ovšem není asociativní ani komutativní. [38], [42], [43], [45], [52]

3 METODY ROZPOZNÁVÁNÍ VÝRAZU OBLIČEJE

Tato kapitola navazuje na kapitolu číslo dvě (Detekce obličeje na základě barvy kůže), po nalezení oblasti obličeje ve vstupním obraze je tato oblast obrazu dále zpracována. V této oblasti se detekují jednotlivé obličejové rysy a poté se určuje výraz obličeje. Rozpoznávání výrazů obličeje lze využít pro analýzu lidských emocí nebo pro zjednodušení komunikace člověka s počítačem pomocí změny výrazu.

Obličejové rysy lze rozdělit do dvou skupin, rysy přechodné (tranzientní) a rysy trvalé (intranziční). Trvalé rysy se jsou v obličeji vždy přítomny, při změně výrazu mohou být deformovány. Mezi trvalé rysy se řadí oční víčka, obočí a ústa. Další skupinu trvalých rysů tvoří například textura tváře nebo rýhy a vrásky, tyto rysy ovlivňují vzhled a výraz obličeje. Mezi přechodné rysy patří různé druhy vrásek a vybouleniny, které nejsou v obličeji přítomny vždy. Tyto rysy se vyskytují převážně v okolí očí a úst. Otevírání a zavírání očí nebo úst výrazně ovlivňuje výraz obličeje.[13]

Proces rozpoznávání výrazu lze rozdělit do dvou částí, které na sebe navazují. První z nich je extrakce příznaků a poté následuje klasifikace výrazů, která těchto příznaků využívá. [1], [2], [3], [13], [48]

3.1 Metody extrakce příznaků

3.1.1 Metody s geometrickým přístupem

Objekty obličeje jsou hledány na základě předpokládané velikosti, tvaru a relativní pozice vůči ostatním objektům a vůči celé oblasti obličeje. Využívá se hranová reprezentace obrazu a prahování. Nejprve jsou detekovány hrany důležitých komponent obrazu a jejich směr, poté je z nich sestaven vektor příznaků. Využívá se hranový detektor a gradientní analýza. Případně může být metoda založena na rozdílu důležitých a nedůležitých komponent šedotónového obrazu, kdy jsou využity Haarovy příznaky, obraz je rozdělen na bloky a pro stanovení vektoru příznaků je využívána metoda Adaboost. Provádí se porovnání středového pixelu dané oblasti s ostatními pixely ze stejné oblasti a na místa okolních pixelů jsou doplněny hodnoty 1 nebo 0 dle výsledků porovnávání. Z takto upravených částí obrazu jsou získány histogramy, které jsou sloučeny s vektorem příznaků vstupního obrazu. Přesnost těchto metod dosahuje až 95%. [1], [2], [3], [11], [44], [48]

3.1.2 Metody založené na šabloně

K nalezení objektů obličeje jsou využívány šablony příslušných objektů. Daný objekt je popsán parametrizovanou šablonou, která udává primární znalost očekávaného tvaru objektu. Využívá se funkce energie, která přiřadí dané části obrazu vhodnou šablonu. Funkce energie je využita pro spojení tvaru vybrané oblasti (peaky, hrany a "údolí" v obraze intenzit) s nejlépe odpovídající šablonou (šablony jsou deformovatelné, aby pokryly danou oblast obrazu). Postupně jsou použity šablony jednotlivých rysů, šablony se posouvají po obraze a provádí se korelace šablony s oblastí obrazu, který je v daném momentě pod šablonou. Oblast s výskytem daného objektu vykazuje po korelaci s příslušnou šablonou maximální hodnoty. Tyto metody selhávají, pokud dojde k výraznější změně vzhledu objektu oproti očekávání, například v případě zavřených očí nebo očí s brýlemi, či v případě otevřených úst. Výhodou této metody je její jednoduchost, nevýhodou je, že šablona rysů musí mít stejnou orientaci a měřítko jako rysy v obraze. Metoda se řadí mezi metody výpočetně náročnější. [1], [3], [13], [44], [48]

3.1.3 Metody založené na jasové informaci barevného obrazu

Metody této skupiny využívají předpokládané barvy jednotlivých objektů, lze využít různé kanály různých barvených prostorů. Pomocí prahování a následné eroze či dilatace jsou získány oblasti odpovídající hledaným objektům. Například ústa jsou hledána jako oblast s vysokými hodnotami v *Cr* kanálu YCbCr prostoru. Problémy mohou vznikat díky umělému nebo nerovnoměrnému osvětlení obličeje. Detekce zavřených očí může být problémová. Metoda je jednoduchá a nepožaduje jiné vstupní znalosti, než prahové hodnoty barevných kanálů pro jednotlivé rysy. Metoda dosahuje přesnosti 85%. Metody z této skupiny byly při řešení této práce využity a jsou podrobněji popsány dále v textu. [1], [2], [13], [48], [61]

3.1.4 Metody založené na vzhledu (Appearance Based Approach)

U metod této skupiny se příznaky obrazu rozumí jakékoli charakteristiky, získané z obrazu. Tyto příznaky nejsou vztahovány ke konkrétním obličejovým rysům (jako např. ústa nebo obočí). Jsou obvykle získávány pomocí filtrace s využitím banky filtrů. Takto určené příznaky nesou důležité informace o obraze, méně důležité informace nejsou uloženy. Efektivní snížení počtu příznaků se zachováním důležitých příznaků může být provedeno pomocí PCA nebo ICA. Výhodou těchto metod je vysoká úspěšnost při rozpoznávání výrazů (98%) a menší počet příznaků obrazu. Nevýhodou jsou nároky na

dobré osvětlení obrazu. [1], [2], [44], [48]

Gaborovy filtry - Základním přístupem pro extrakci příznaků je využití Gaborových filtrů. Vstupní obrazy obsahující převážně oblast obličeje jsou zpracovány pomocí afinní transformace k normalizovaným obličejům. Hodnocení obrazů je provedeno s pomocí jednotlivých Gaborových filtrů, což jsou 2D lineární pásmové propusti s různými charakteristikami. Filtry jsou využity k odlišení jednotlivých výrazů obličeje ve výstupních obrazech. Banka Gaborových filtrů obsahuje několik 2D filtrů s různými orientacemi a frekvencemi sinusoidálního signálu a s různými směrodatnými odchylkami, jedná se o komplexní sinusoidální signál modulovaný Gaussovou funkcí. Výstupem jsou odezvy na jednotlivé Gaborovy filtry, které jsou dále podvzorkovány a tvoří vektor příznaků vstupního obrazu.

Výhodou metody obecně je její invariance k rotaci obličeje v obraze, k jeho velikosti a umístění v obraze, dále odolnost proti šumu v obraze a proti vlivu nerovnoměrného osvětlení. Tato metoda je úspěšná i při výskytu brýlí nebo vousů v obličeji, při zavřených nebo částečně zakrytých očích a při různých pozicích hlavy. Přesnost metody dosahuje až 93%. [1], [5], [7], [13], [18], [48]

Diskrétní kosinová transformace (DCT) - Dalším zástupcem metod založených na vzhledu je diskrétní kosinová transformace, jejíž koeficienty vzniklé aplikací DCT na obraz jsou uchovány v zig-zag pozici a jsou použity ke konverzi 2D DCT obrazové matice na vektor příznaků. Jako klasifikátor se používá AdaBoost a metoda rozpoznání výrazu dosahuje přesnosti 76%. [1], [2], [48]

3.1.5 Jiné možné rozdělení metod extrakce příznaků

Kromě uvedeného rozdělení metod pro získání příznaků z obrazu existují i jiná možná rozdělení. Některá z nich jsou uvedena v následujících odstavcích.

Metody s lokálním přístupem a holistické metody. Holistické metody zpracovávají obraz jako celek, zatímco metody s lokálním přístupem zpracovávají obraz po částech a pracují hlavně s oblastmi, ve kterých dochází ke změnám při změně výrazu. Holistické metody jsou dobré pro rozpoznávání obvyklých výrazů, lokální metody jsou schopné detekovat detailní změny v menších oblastech. [13]

Deformační metody a metody založené na pohybu. Metody založené na pohybu se přímo zaměřují na změny tváře způsobené změnou výrazu obličeje. Metody založené na deformaci využívají obrazu neutrální tváře nebo model tváře a s jeho pomocí extrahují rysy obličeje, které jsou důležité pro rozpoznávání výrazů (vrásky způsobené stářím se v tomto případě neberou v potaz). Deformací se rozumí změny tvaru nebo textury dané oblasti. Metody založené na deformaci mohou být použity pouze na jeden snímek nebo i na sekvenci snímků, metody založené na pohybu lze z jejich podstaty

použít pouze na sekvenci snímků. [13]

Metody založené na obrazové informaci a metody využívající model obličeje. Metody založené na obrazové informaci nevyžadují rozsáhlejší vstupní informace o objektu zájmu, výhodou je jejich rychlost a jednoduchost. Modely obličeje popisují strukturu tváře a mohou být 2D nebo 3D. Modely obličeje nebo hlavy jsou konstruovány manuálně, což lze chápat jako nevýhodu těchto metod. Další nevýhodou využívání modelů tváře je skutečnost, že vyžadují komplexní mapování, které bývá výpočetně náročné. [13]

3.2 Klasifikátory výrazů obličeje

Klasifikace výrazů je poslední krok v automatickém rozpoznávání výrazů obličeje. Na základě získaných příznaků (popsáno v kap 3.1) klasifikátory rozhodují, o jaké výrazy se na vstupních obrazech jedná.

3.2.1 Skryté Markovské modely (HMM)

HMM je markovský proces, který obsahuje skryté neznámé parametry, a umožňuje efektivně popsat statistický model náhodného signálu. HMM jsou úspěšně využívány v rozpoznávání řeči a v poslední době stoupá i jejich význam v rozpoznávání tváří a jejich výrazů. HMM lze použít pro klasifikaci základních výrazů, ale HMM umožňují i klasifikovat emoce vyšších úrovní, jako jsou nejistota, zaujatý výraz, nesouhlas, povzbuzující a odrazující výraz, se znalostí základních emocí, mezi které patří radost, překvapení, neutrální výraz nebo smutek. Emoce ze vstupního obrazu je porovnávána s výrazy obličeje uloženými v databázi. HMM je možné použít i pro rozpoznávání výrazu z videa, rozpoznávání probíhá v reálném čase. Výhodou HMM je odolnost vůči většímu natočení obličeje vůči kameře nebo fotoaparátu a v případě přidání nových vzorků není třeba znovu provést učení algoritmu. [20], [48], [61]

3.2.2 Neuronová síť

Neuronová síť umožňuje klasifikaci výrazů přímo z obrazů obličejů, často se ale používá v kombinaci s některou metodou pro získávání příznaků, klasifikace probíhá ve dvou stupních. Prvnímu stupni předchází předzpracování obrazu, které má zajistit snížení časové náročnosti algoritmu a zvýšení kvality obrazu. V prvním stupni klasifikace jsou získány příznaky obrazu například pomocí Gaborových filtrů s různou orientací a různou frekvencí (např. banka 20 filtrů). Kromě Gaborových filtrů lze využít

i metod PCA nebo ICA pro redukci počtu příznaků. Ve druhém stupni dochází ke klasifikaci vektoru příznaků pomocí neuronové sítě.

Neuronovou síť lze popsat jako nekontrolovanou statistickou analytickou metodu, která provádí značnou redukci dimenzionality, což zjednodušuje a zlepšuje následující klasifikaci. Pro větší škálu výrazů je obtížné neuronovou síť natrénovat. Problémem je velmi vysoký počet možných kombinací změny tváře.

Neuronová síť je trénována pomocí algoritmu zpětného šíření chyby (EBP, Error Back Propagation).

EBP je zobecněním Widrow-Hoffova pravidla pro vícevrstvé sítě. Algoritmus je založen na využití diferencovatelných funkcí, jako je například sigmoida, které nahradí funkci prahu používanou v jednoduchých lineárních sítích. Algoritmus je založen na zpětném šíření chyby z výstupního neuronu přes vnitřní/skryté vrstvy k vstupu sítě, chyba se šíří stejným způsobem, jako je přenášen signál normálních operací od vstupu přes vnitřní/skryté vrstvy až k výstupu, pouze v opačném směru. Neurony na váhovanou sumu vstupů používají aktivační funkci, nejčastěji používají sigmoidální funkci, která je uvedena v následující rovnici [10]:

$$f(x) = \frac{1}{1+e^x}, \quad (4.1.)$$

kde $f(x)$ je sigmoidální funkce a x je vstupní proměnná. [10]

Počet neuronů ve vstupní vrstvě sítě je roven počtu prvků ve vektoru příznaků. Počet neuronů ve výstupní vrstvě sítě udává počet rozpoznatelných výrazů. Pro fázi učení musí mít každý vstupní obraz přiřazenou hodnotu z výstupního vektoru. Síť může mít libovolný počet skrytých vrstev neuronů umístěných mezi vstupní a výstupní vrstvou. Každý neuron dané vrstvy je propojen se všemi neurony následující vrstvy.

Učení sítě probíhá na trénovací množině dat, která obsahuje vstupy a výstupy sítě. Na vstup sítě jsou postupně předkládány jednotlivé vstupy trénovací množiny dat. Výstup sítě je získán postupem vstupního signálu sítě od vstupu přes skryté vrstvy směrem k výstupu s využitím aktuálních hodnot vah sítě, jedná se tedy o dopřednou síť (forward propagation). Poté je vypočtena chyba výstupu se znalostí požadovaných výstupů jako součet kvadrátů chyb z každého výstupního neuronu. Tato chyba se zpětně šíří skrz síť a vede ke změně nastavení vah. Tento postup se opakuje pro všechny vstupy trénovací množiny dat. Pokud během učení sítě dojde k poklesu chyby pod stanovenou mez, je učení v daném okamžiku ukončeno. Učení spočívá v minimalizaci celkové kvadratické chyby objevující se na výstupních neuronech během celé doby učení sítě, kdy jsou na vstup vkládána data z trénovací množiny. Minimalizace je prováděna pomocí gradientní metody, kdy jsou počítány derivace výsledné kvadratické chyby pro jejich váhy. Proto jsou požadovány spojitě derivace funkce neuronu.

Klasifikace pomocí neuronových sítí dosahuje přesnosti 84%. [1], [10], [13], [20], [48]

3.2.3 Support Vector Machine (SVM)

Vstupní obrazy jsou upraveny na určitou vstupní velikost a dále filtrovány bankou Gaborových filtrů. Filtrované obrazy vstupují do klasifikátoru, který zakóduje výraz obličeje v obrazu do různých dimenzí. Příznaky obrazu jsou vybrány z Gaborových filtrů pomocí metody AdaBoost a následuje trénování metodou SVM. Tento algoritmus je možné použít i na video obsahující obličeje s určitými výrazy. Přesnost metody se pohybuje okolo 93%. [1], [20], [48]

3.2.4 AdaBoost

Metoda AdaBoost (zkratka z Adaptive Boosting) je klasifikační metoda, která může být využita v široké škále oblastí klasifikace. Pro detekci výrazu tváře může být využita v kombinaci s příznaky získanými pomocí metod Color Space Transformation, Connected Component Labeling Technique a Face Recognition Verification jako segmentace tváře a detekce zorniček. Obraz tváře je poté klasifikován do jedné ze sedmi kategorií: radost, smutek, překvapení, rozzlobenost/vzteky, strach a odpor. Přesnost tohoto algoritmu dosahuje až 90%. [20], [48], [61]

3.2.5 Algoritmus nejbližších sousedů (K-Nearest-Neighbor, KNN)

Jedná se o relativně jednoduchou metodu pro rozpoznávání vzorů, která dosahuje srovnatelných výsledků s jinými metodami. Tento algoritmus hledá K nejbližších sousedů aktuálního vzorku, a následně tento vzorek zařadí do skupiny, do níž patří největší počet z K nejbližších sousedů. Jedná se o metodu učení s učitelem, učení probíhá na trénovací množině, klasifikace dalších prvků probíhá v prostoru trénovací množiny, do které je nový prvek vložen, a k němu jsou hledáni nejbližší sousedé. [48], [65]

3.2.6 Eigenface (eigenpicture, eigenvector)

Eigenface je jedním z nejpropracovanějších přístupů k rozpoznávání obličejů, lze ho však využít i k rozpoznávání výrazů. Pro efektivní reprezentaci obrazů je použita analýza hlavních komponent. Hlavní princip metody eigenface je ten, že jakýkoli obličej lze přibližně rekonstruovat pomocí malé množiny vah pro každý obličej a ze standardního obličeje (eigenface). Váhy pro každý obličej jsou získány promítnutím tohoto obličeje přes standardní obličej (eigenface). Každý obličej může být přesně reprezentován lineární kombinací standardních obličejů. Nejlepších M standardních

obličejů tvoří M dimenzionální prostor tváří ("face space").

Z matematického pohledu jsou standardní obličejové hlavními komponentami rozložení obličejů, nebo je možné na ně pohlížet jako na vlastní vektory kovarianční matice souboru obrazů obličejů. Metoda je rychlá a odolná vůči různým typům osvětlení obličejů v obrazech.

Stejným způsobem může být podle Lambertiana a spol. [39] metoda použita na obličejové rysy s využitím eigenfeatures, které odpovídají očím, ústům a nosu (eigeneyes, eigenmouth, eigennose). Tato metoda je méně citlivá ke změnám než klasická metoda eigenfaces, dosahuje přesnosti okolo 95%. [2], [20]

3.2.7 Fisherfaces

Tato metoda je navržena s použitím analýzy hlavních komponent a Fisherovy lineární diskriminační analýzy a metoda vytváří podprostorovou projekční matici, která je velmi podobná prostoru vznikajícímu použitím metody Eigenface. Fisherfaces využívá informací uvnitř tříd (minimalizuje odchylky uvnitř skupin a maximalizuje separaci skupin navzájem). Výhodou oproti metodě Eigenface je, že Fisherfaces zvládne překonat obtížnosti klasifikace spojené s rozdíly osvětlení a pózami hlavy v obraze. [2], [20]

3.2.8 Bayesovská klasifikace (Bayesian classification)

Bayesovská klasifikace vychází z Bayesovy věty pro podmíněnou pravděpodobnost, nicméně tato metoda se často využívá v souvislosti se strojovým učením. Bayesovské metody umožňují klasifikaci do tříd s určitou pravděpodobností. Na rozdíl o jiných algoritmech učení se v tomto případě neprovádí prohledávání prostoru hypotéz, stačí pouze spočítat příslušné pravděpodobnosti z četností výskytů hodnot jednotlivých atributů. [13], [61]

3.2.9 Metody redukující vektor příznaků

Metody extrakce příznaků často produkují velké množství příznaků, které ale nemusí nést dostatečnou informaci ke klasifikaci obličejů do výstupních tříd a jsou tedy redundantní. Pro detekci a eliminaci redundantních příznaků je možné využít některou z následujících metod. Takto redukovaný prostor příznaků umožní rychlejší činnost klasifikátoru s minimálním snížením úspěšnosti klasifikace. Mezi takovéto metody, které klasifikaci předchází, patří analýza hlavních komponent (PCA), analýza nezávislých komponent (ICA), či lineární diskriminační analýza (LDA). [1], [20], [48]

Analýza hlavních komponent (PCA) - Cílem PCA je zjednodušení popisu lineárně závislých znaků a redukce počtu příznaků bez velké ztráty informace užitím

pouze několika hlavních komponent. Jde o přepis vstupních dat do jiné souřadné soustavy, nová souřadná soustava je tvořena pomocí matice vlastních vektorů kovarianční matice, vybírají se komponenty například s největším rozptylem.

Analýza nezávislých komponent (ICA) - ICA je metoda separace dat nebo signálů, které byly smíchány lineární kombinací původních signálů. To umožňuje v původních datech odhalit skryté informace, které nejsou ve smíchaných datech patrné. K lineární separaci jednotlivých složek metoda používá kritérium statistické nezávislosti. Jedná se tedy spíše o nalezení skrytých informací. Na jejich základě mohou být pomocí PCA redundantní data odstraněna.

Lineární diskriminační analýza (LDA) - LDA představuje výpočet diskriminačního vektoru založeného na Fisherově kritériu, pracuje se ve dvou stupních. V prvním jsou homogenní regiony obrazu tváře roztříděny do skupin na základě geometrických charakteristik. Průměrná jasová hodnota (šedotónového obrazu) je použita pro pixely uvnitř jednotlivých regionů. Díky tomu je zredukován vektor příznaků. V druhém případě určuje vektor příznaků osy diskriminační projekce založené na LDA. Vektory jsou potom shlukovány (tříděny) do skupin pomocí K-mean clusteringu použitého na každý prvek.

[13], [20], [61]

4 DETEKCE OBLIČEJE A OBLIČEJOVÝCH RYSŮ

Pro realizaci praktické části diplomové práce byla pro detekci obličeje a obličejových rysů zvolena metoda založená na neměnných rysech obličeje, v tomto případě na barvě kůže. Je to implementačně jednoduchá a výpočetně rychlá metoda, kterou není třeba trénovat. Detekce obličeje v obraze by měla zvládnout nalezení různých lidských obličejů různé velikosti na různých pozicích v obraze. Různá orientace obličeje, jeho výraz ani rozdíl mezi barvami kůže různých ras by neměly mít na detekci vliv, stejně tak jako komplexní pozadí, nebo kvalita obrazu. Detekci obličejů je vhodné provádět na barveném obraze, neboť barva je vhodný parametr pro nalezení regionů kůže.

Pro detekci obličejů v obraze byla zvolena barva kůže jako ukazatel, zda se v daném místě obličej nachází. Každý pixel obrazu je porovnán s podmínkami pro barvu kůže a je označen jako pixel s barvou kůže nebo jako pixel jiné barvy. Pro klasifikaci pixelů není třeba znát geometrické vlastnosti ani texturu kůže. V obraze se mohou kromě obličeje vyskytovat i jiné části lidského těla s barvou kůže. Proto jsou detekované oblasti kůže dále zpracovávány pomocí morfologických operací, které odstraní jiné detekované objekty než obličej včetně chybně detekovaných oblastí, které neodpovídají barvě kůže.

Odlišení pixelů odpovídajících kůži od ostatních pixelů je možno provádět v různých barevných prostorech, které jsou popsány v kapitole 1.1. Detekce pouze na základě kanálů popisujících barvu bez vlivu osvětlení je pro detekci různých barev kůže efektivnější než detekce využívající kanály ovlivněné osvětlením. Příkladem mohou být kanály H , S prostoru HSV, kanál zaznamenávající osvětlení (u HSV prostoru je to kanál V) není při detekci brán v potaz. Lze využít i RGB prostor, kde je osvětlení přítomno ve všech třech kanálech, ale před samotnou detekcí je vhodné vliv osvětlení odstranit nebo zkombinovat detekci v RGB prostoru s jiným prostorem, kde osvětlení neovlivňuje všechny kanály. [46], [50]

4.1 Detekce obličeje podle barvy kůže v barevných prostorech RGB, HSV a YCbCr

V této práci byla vybrána detekce obličeje podle barvy kůže s využitím třech barevných prostorů: RGB, HSV a YCbCr. Kombinací více barevných prostorů a sloučením detekce v každém prostoru je možné dosáhnout lepších výsledků, než při využití pouze jednoho barevného prostoru.

Podmínky pro jednotlivé kanály barevných prostorů je možné odhadnout z histogramu obrazu. Po otestování podmínek publikovaných v různých dostupných zdrojích byly vybrány a použity podmínky z článku [50], dílčí podmínky jsou spojeny logickým AND. V RGB prostoru není oblast barvy kůže dobře odlišitelná, z histogramu vyplývá, že barva kůže je v kanálech RGB rozložena v širokém rozpětí hodnot. Jelikož v RGB prostoru osvětlení ovlivňuje všechny tři kanály, byly použity dvě kategorie podmínek, první pro rovnoměrné osvětlení denním světlem a druhá pro umělé nebo nerovnoměrné osvětlení. Tyto dvě kategorie jsou spojeny logickým nebo (OR). (RGB prostor předpokládá hodnoty v rozsahu 0-255). V prostoru HSV umožňuje kanál *H* jasně odlišit oblasti kůže od ostatních oblastí. Z YCbCr prostoru byly využity kanály *Cb* a *Cr* popisující barvu pixelu. Všechny použité podmínky jsou uvedeny v tabulce 4-1.

Tabulka 4-1 Přehled podmínek pro barvu kůže pro jednotlivé kanály prostorů RGB, HSV a YCbCr [50]

Barevný prostor:	Podmínka:
RGB za rovnoměrného denního světla	$R > 50$
	$G > 40$
	$B > 20$
	$\max(R, G, B) - \min(R, G, B) > 10$
	$ R - G \geq 11$
	$R > G$
	$R > B$
RGB za umělého nebo nerovnoměrného osvětlení	$R > 220$
	$G > 210$
	$B > 170$
	$ R - G \leq 15$
	$R > B$
	$G > B$
HSV	$H \geq 0$
	$H \leq 50$
	$S \geq 0.1$
	$S \leq 0.9$
YCbCr	$Cb \geq 60$
	$Cb \leq 130$
	$Cr \geq 130$
	$Cr \leq 165$

Algoritmus byl implementován v prostředí Matlab. Dílčí podmínky byly spojeny logickými operátory, jak již bylo uvedeno výše. Výstupem detekce byl binární dvojrozměrný obraz o stejné velikosti jako vstupní obraz. Hodnoty rovné nule ve výstupním obraze udávají oblasti bez výskytu barvy kůže, zatímco hodnoty rovné jedné označují oblasti, kde barva pixelů odpovídá barvě kůže. [24], [36], [46], [50]

4.2 Aplikace morfologických operací

Binární obraz s vyznačenými oblastmi kůže byl dále zpracováván pomocí morfologických operací, aby byl obraz pročištěn – malé a chybně detekované oblasti budou odstraněny a oblast obličeje jako celek se stane více kompaktní.

Nejprve byl pro každou oblast kůže s možným výskytem obličeje spočítán počet pixelů oblasti. Pokud byla oblast tvořena méně pixely než stanovená mez, bylo rozhodnuto, že tato oblast nepředstavuje oblast obličeje a oblast byla odstraněna. Hodnota meze byla nastavena na 1500 pro obrazy s výškou 200 pixelů. Předpokládá se, že oblast s méně než 1500 pixely nemůže představovat oblast obličeje. Tato hodnota byla zvolena experimentálně a její hodnota se může měnit s různými vstupními obrazy.

Nalezení oblastí odpovídajících barvě kůže v binárním obraze bylo provedeno v prostředí Matlab pomocí funkce `bwconncomp`. Počet pixelů jednotlivých oblastí byl určen pomocí funkce `cellfun`, a následně byl počet pixelů každé oblasti porovnán s předem stanovenou mezí a případně byla tato oblast odstraněna.

Malé oblasti, které neodpovídají barvě kůže a nacházejí se uvnitř oblastí kůže, byly sjednoceny s oblastí kůže pomocí funkce `imfill`. Mezi takové oblasti patří například oblast očí nebo obočí.

K pročištění obrysů oblastí kůže bylo využito morfologických operací eroze a dilatace. Nejprve byly pomocí funkce `imerode` (zajišťuje erozi obrazů) odstraněny ostrůvky oblastí kůže vybíhající do oblastí, které barvě kůže neodpovídají. Dále následovalo opět odstranění oblastí kůže s počtem pixelů menším než stanovená mez. Tato operace je využita k odstranění malých oblastí kůže, které se v důsledku eroze oddělily od oblasti obličeje. V důsledku eroze došlo také ke zmenšení zachovaných oblastí kůže, proto následovalo použití morfologické operace dilatace pomocí funkce `imdilate`, která oblasti opět rozšíří.

Pokud se v obraze nachází více potenciálních oblastí obličeje, dojde k jejich očíslování pomocí funkce `bwlabel`. Dále se postupně zpracovává každá z těchto oblastí zvlášť.

Dále byl zkoumán tvar oblastí s barvou kůže, pomocí funkce `regionprops` byly získány parametry obdélníku, který oblast z venčí ohraničuje, a excentricita. Pro poměr stran obdélníku ohraničující danou oblast platí, že pokud je poměr výšky k šířce oblasti v rozmezí 0,6 až 1,4, je daná oblast považována za oblast obličeje [60_10 c]. Jelikož je oblast obličeje obvykle neoddělitelně spojená s oblastí krku, bylo rozmezí hodnot pozměněno na 0,6 až 2,1. Oblasti, jejichž poměr výšky k šířce nesplnil toto rozmezí, nebyly považovány za obličej a dále se s nimi nepracovalo. Excentricita udává poměr vzdálenosti ohnisek a délkou hlavní osy elipsy, která má stejný druhý moment jako daná oblast, hodnota se pohybuje v rozmezí od 0 do 1. Oblasti, jejichž excentricita byla menší než hodnota 0,25, nebyly považovány na oblasti obličeje [66].

Výsledné oblasti obličejů v binárním obraze byly ohraničeny obdélníky a poté na pozicích těchto obdélníků byly v barevném obraze hledány jednotlivé objekty obličeje (oči, nos a ústa), které byly využity pro zpřesnění ořezu oblasti obličeje tak, aby se výsledné obrazy co nejvíce podobaly ořezaným obrazům z databáze AR. Algoritmus tedy umožňuje nalezení více obličejů v jednom obraze. Jelikož se ale předpokládá vstup obrazů z databáze obličejů a jim podobné, byly další úpravy a hledání objektů obličeje implementovány pro jednu oblast obličeje v každém obraze.

Obrazy oblastí obličejů byly v malých oblastech, kde se nenachází obličej, doplněny o černé pozadí. Tím se eliminuje vliv pozadí nebo jiných objektů (například vlasů), které by mohly způsobit nepřesnosti následné detekce rysů obličeje. Tímto se obrazy odlišují od manuálně ořezaných obrazů AR databáze, které mají bílé pozadí. Ovšem tento rozdíl se v rozpoznávání výrazů neprojeví, neboť výsledky filtrace obrazu bankou filtrů dávají v oblasti mimo obličej vždy stejné výsledky, bez ohledu na barvu pozadí (viz obrázek 5-5 a 5-7).

[24], [36], [50], [53], [58]

4.3 Detekce charakteristických rysů obličeje

4.3.1 Detekce středů očí

Pro nalezení očí bylo zvoleno několik odlišných metod, které využívají informace z jednotlivých kanálů YCbCr prostoru. Na obrazech, které byly pořízeny při dobrém osvětlení, a které mají dobré vyvážení barev, fungují všechny metody. U obrazů pořízených v různorodém prostředí za různého osvětlení a při různé sytosti a vyvážení barev může dojít k selhání některé z metod. Proto bylo implementováno více různých metod a případné selhání některé z nich bylo kompenzováno výsledky ostatních metod.

Oblasti očí mohou být podle [44] indikovány vysokými hodnotami kanálu C_b a nízkými hodnotami kanálu C_r barevného prostoru YCbCr. První metoda pro nalezení očí byla založena na hledání oblastí, v nichž hodnoty kanálu C_b jsou o určitý násobek větší než hodnoty kanálu C_r . Takto vznikl binární obraz s vyznačenými oblastmi, které splňují uvedenou podmínku. Z binárního obrazu byly odstraněny oblasti, které byly menší velikosti než stanovený práh, čímž byly odstraněny oblasti, které neodpovídaly pozici očí. Ve zbývajících oblastech byl určen střed oblasti jako pozice středu oka. [44]

Druhá metoda pro nalezení oblastí očí v obraze byla založena na hledání maximálních hodnot v mapě získané výpočtem z hodnot kanálů C_b a C_r barevného prostoru YCbCr. Metoda využívá faktu, že barva zornice a bělma je u většiny osob velmi podobná. Pro výpočet binární mapy očí byla použita následující rovnice [58]

$$M = \frac{1}{3} C_b^2 + (256 - C_b)^2 + \frac{C_b}{C_r}, \quad (4.1)$$

kde M představuje binární mapu, C_b a C_r jsou kanály barevného prostoru YCbCr (s rozsahem hodnot od 0 do 255). Jelikož barevná skladba rtů splňuje podobné podmínky jako barevná skladba očí, mohly by se rty v mapě projevit vysokými hodnotami. Proto byla mapa očí počítána pouze pro oblast nad úroveň špičky nosu. Dvě maximální hodnoty mapy, které neleží ve své blízkosti, odpovídají nalezeným pozicím očí. [58]

Třetí metoda hledání oblastí očí využívala binární mapu, která byla vytvořena na základě faktu, že oblast očí udává vysoké hodnoty barevné složky C_b prostoru YCbCr [7]. V binární mapě byly odstraněny oblasti bílých pixelů o velikosti menší než stanovená mez. Zbýlé oblasti bílých pixelů byly detekovány jako oči. U osob s tmavší barvou obočí může dojít i k detekci obočí. [7]

Pro zpřesnění detekce očí byla využita funkce `CascadeObjectDetector` z toolboxu Computer Vision System Toolbox v prostředí Matlab. Detektor využívá Viola-Jones algoritmu a umožňuje detekovat lidskou tvář, oči, nos, ústa či vrchní polovinu lidského těla. Následně byl výstup detektoru použit jako vstup funkce `step`, která na výstupu vrací pozice rohů obdélníku, který oblast oka ohraničuje. Poté byly vypočteny středy očí. Tato metoda dává dobré výsledky i v přítomnosti brýlí v obličeji, naopak ale selhává, pokud jsou oči zavřené.

Hledání očí probíhalo v oblasti obličeje od horního okraje obrazu po úroveň špičky nosu (v případě, že nos byl detekován, v opačném případě probíhala detekce v celé oblasti obrazu). Pozice očí získané různými metodami byly vzájemně porovnány a pokud se souřadnice očí získané některou metodou odlišovaly od ostatních o více než stanovenou toleranci, nebyly tyto hodnoty pro další výpočty brány v potaz. Pokud se všechny metody od sebe výrazně odlišují, jsou pro další výpočty použity souřadnice očí získané pomocí algoritmu Viola-Jones, který se ukázal jako nejrobustnější. U obrazů pořízených za dobrých podmínek osvětlení obličeje fungují všechny uvedené metody a

souřadnice očí jednotlivých metod se liší maximálně o pár pixelů, znázorněno je to na obrázku 4-1. U obrazů s horší kvalitou osvětlení může některá z metod detekce očí selhat (viz obrázek 4-2).

4.3.2 Detekce středu úst

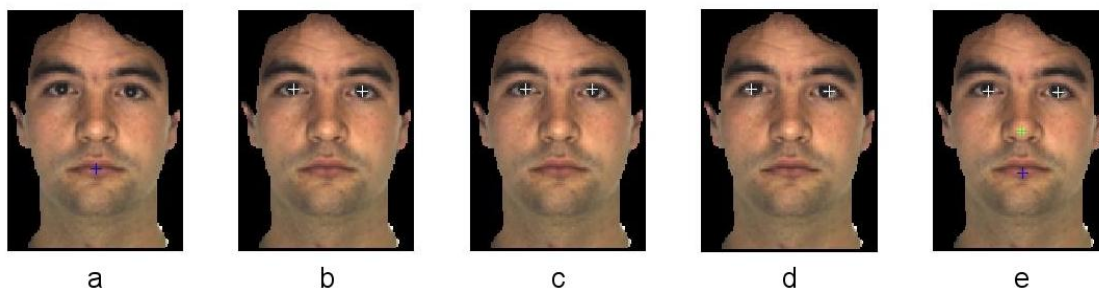
K nalezení úst byl využit normalizovaný kanál zelené barvy prostoru RGB. Podle [58] se ústa nachází v místech obrazu, kde je hodnota normalizované zelené barvy menší než stanovená mez (mez byla experimentálně určena na 0,25). Výsledek detekce úst je zobrazen na obrázku 4-1 a) na otce z AR databáze a na obrázku 4-2 a) na fotce z vlastní databáze.

Další metoda pro nalezení úst byla založena na barevné FLD transformaci, metoda předpokládá, že oblast úst představuje vysoké zastoupení červené barvy a nízké zastoupení modré barvy RGB prostoru. Tato metoda ovšem nepostihuje všechny osoby na fotografiích, barevná skladba oblasti úst se u různých osob liší. Proto nebyla tato metoda použita.

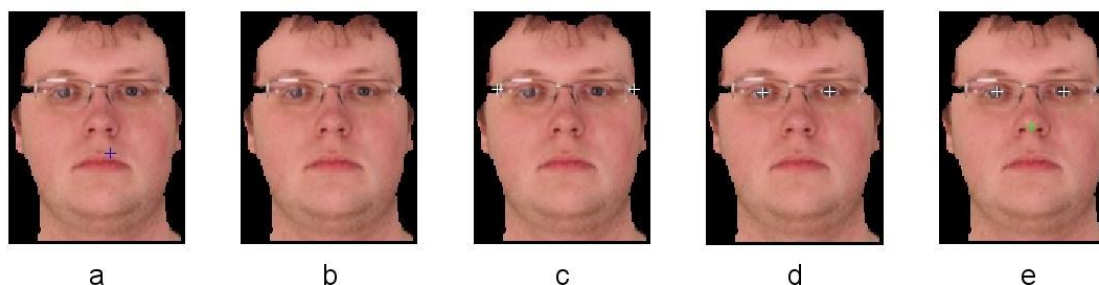
V případě, že prahováním normalizovaného kanálu zelené barvy RGB prostoru nebyla ústa nalezena, byla pro detekci úst využita funkce `CascadeObjectDetector` z toolboxu programu Matlab, jak bylo popsáno v předchozích odstavcích, nyní však funkce detekovala oblast úst. Pomocí funkce `step` byla detekovaná oblast ohraničena a následně vypočtena pozice středu úst. Detekovaná ústa jsou zobrazena na obrázku 4-1 e). [7], [58]

4.3.3 Detekce špičky nosu

Pro nalezení nosu byla využita funkce `CascadeObjectDetector` z toolboxu Computer Vision System Toolbox v prostředí Matlab. Detektor využívá Viola-Jones algoritmu a umožňuje detekovat lidskou tvář, oči, nos, ústa či vrchní polovinu lidského těla. Následně byl výstup detektoru použit jako vstup funkce `step`, která na výstupu vrací pozice rohů obdélníku, který oblast nosu ohraničuje. Následně byla vypočtena pozice špičky nosu jako střed ohraničené oblasti. Detekce nosu je znázorněna na obrázku 4-1 e) zelenou barvou. [36]



Obrázek 4-1 Ukázka zdařilé detekce úst, očí a nosu. a) detekce úst v prostoru RGB, b) detekce očí první metodou, c) detekce očí druhou metodou, d) detekce očí třetí metodou, e) detekce očí, nosu a úst pomocí Viola-Jones algoritmu



Obrázek 4-2 Ukázka méně zdařilé detekce úst, očí a nosu. a) detekce úst v prostoru RGB, b) detekce očí první metodou, c) detekce očí druhou metodou, d) detekce očí třetí metodou, e) detekce očí, nosu a úst pomocí Viola-Jones algoritmu

Na obrázku 4-1 je ukázka zdařilé detekce rysů obličeje v obrázku z AR databáze. Různé metody pro nalezení středu očí se liší v řádech několika pixelů. Na obrázku 4-2 je ukázka méně zdařilé detekce rysů na obraze z vlastní databáze. Ústa na obrázku 4-2 a) jsou detekovaná nepřesně, detekován byl spíše horní okraj rtů. Obrázek 4-2 b) ukazuje, že oči hledané první metodou nebyly nalezeny, druhá metoda detekce očí (c) označila jako středy očí nožičky brýlí, třetí metoda detekce očí (d) detekuje oči správně. Detekce úst pomocí algoritmu Viola-Jones v tomto případě selhala.

4.3.4 Automatický ořez obličeje

Se znalostí pozic středu očí, úst a špičky nosu byla vypočítána oblast, která obsahuje pouze oblast obličeje, shodně s obrazy oblastí obličeje z databáze AR. Kritéria výběru oblasti obličeje byla následující: spodní hranicí výběru byla hrana brady, horní hranicí výběru byl přechod oblasti čela v začátek oblasti vlasů nad čelem, poměr stran vybrané oblasti šířka ku výšce byla 7:10, střed špičky nosu se nachází v polovině šířky vybrané oblasti, vybraná oblast by v ideálním případě neměla obsahovat uši.

Hranice oblasti obličeje byly vypočítány ze souřadnic středu úst, očí a špičky nosu následovně. Souřadnice nosu byly použity pouze tehdy, pokud souřadnice očí nebo souřadnice úst nebyly určeny dostatečně přesně (případně pokud nebyly nalezeny vůbec). Pokud by nebyly nalezeny souřadnice dvou ze tří objektů (očí, nos, ústa), nemohl by být automatický ořez proveden. U dvaceti vybraných obrazů různých osob byly vypočítány poměry vzdáleností mezi jednotlivými význačnými body obličeje. Střední hodnoty těchto poměrů vzdáleností uvádí následující tabulka.

Tabulka 4-2 Střední hodnoty poměrů vzdáleností mezi význačnými body obličeje

Objekty použité k výpočtu hranice oblasti obličeje	Poměr vzdáleností mezi význačnými body obličeje	
	Hranice hrany brady	Hranice čela
ústa, střed mezi očima	$\frac{\text{vzálenost ústa} - \text{brada}}{\text{vzdálenost očí} - \text{ústa}} = 0,7$	$\frac{\text{vzálenost čelo} - \text{oči}}{\text{vzdálenost očí} - \text{ústa}} = 1,2$
ústa, nos	$\frac{\text{vzálenost ústa} - \text{brada}}{\text{vzdálenost nos} - \text{ústa}} = 1,4$	$\frac{\text{vzálenost čelo} - \text{nose}}{\text{vzdálenost nos} - \text{ústa}} = 3,3$
nos, střed mezi očima	$\frac{\text{vzálenost nos} - \text{brada}}{\text{vzdálenost očí} - \text{nose}} = 2,2$	$\frac{\text{vzálenost čelo} - \text{oči}}{\text{vzdálenost očí} - \text{nose}} = 2,1$

Pomocí poměrů vzdáleností význačných bodů obličeje byly vypočteny horní a spodní hranice oblasti obličeje. Pokud se tyto hranice nacházely uvnitř obrazu, mohl být obraz podle těchto hranic oříznut. Boční hranice oblasti obličeje byly dopočítány tak, aby výsledný obraz splňoval výše uvedený poměr šířky ku výšce 7:10. Střed špičky nosu určuje střed šířky oblasti obličeje.

V této části bylo důležité čelní postavení obličeje a vzpřímené držení hlavy. Natočení a úhel mohou při automatickém ořezu způsobovat nepřesnosti. Úklon hlavy (přiklonění hlavy k jednomu rameni) byl u obrazů, u nichž byl patrný, korigován ručně. Algoritmus by bylo možné doplnit o automatickou korekci úklonu hlavy. Korekce by mohla být počítána na základě znalostí pozice středu úst a středu očí. Úhel, pod kterým je hlava odkloněna od svislé osy, je možné vypočítat ze znalosti vzdálenosti středu úst od středu mezi očima a rozdíl souřadnic těchto bodů ve svislém směru (výpočet pomocí funkce kosinus). Automatická korekce úklonu však klade značné nároky na přesnost detekce středu očí a úst. Jelikož někdy dochází k nepřesným detekcím těchto rysů a obrazy z databází mají malý nebo žádný úklon hlavy, nebyla automatická korekce do implementovaného algoritmu zahrnuta. Kromě automatického ořezu může mít úklon hlavy negativní vliv i na výpočet Gaborových příznaků obrazu. Jak je uvedeno v

kapitole 5.2, byly Gaborovy filtry konstruovány pro různé úhly šíření prostorových vln, čímž byla zajištěna separace příznaků vynikajících v různých směrech. Úklon hlavy pozměňuje natočení důležitých rysů obličeje a to může mít vliv na výsledek Gaborových příznaků.

Navržený výběr oblasti obličeje je v obslužném skriptu zobrazen uživateli a uživatel má na výběr, zda tento výběr použije, nebo zda bude chtít oblast obličeje vybrat ručně.

Ruční výběr oblasti obličeje v obraze byl realizován pomocí funkce `imrect`. Vybraná oblast je označena obdélníkem, který je možno po obraze posouvat a měnit jeho velikost, poměr stran obdélníku však zůstává zachován, aby vybraná oblast odpovídala parametrům manuálně ořezaných obrazů z AR databáze. Ruční výběr zadaný uživatelem pomocí umístění obdélníku s šedým ohraničením přes oblast obličeje je znázorněn na obrázku 4-3.

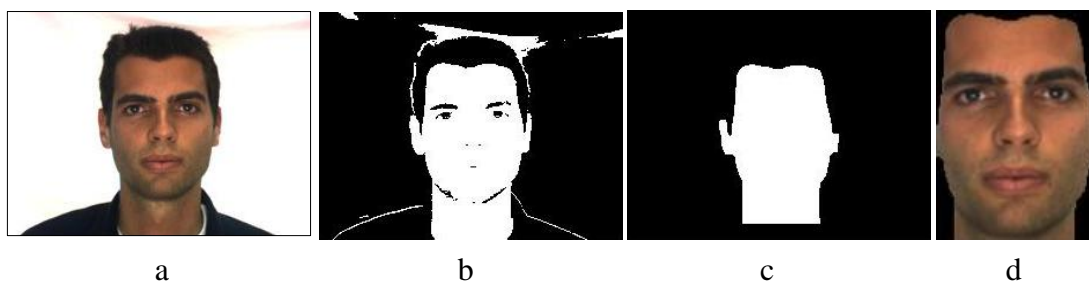


Obrázek 4-3 Zobrazení ručního výběru oblasti obličeje pomocí funkce `imrect`

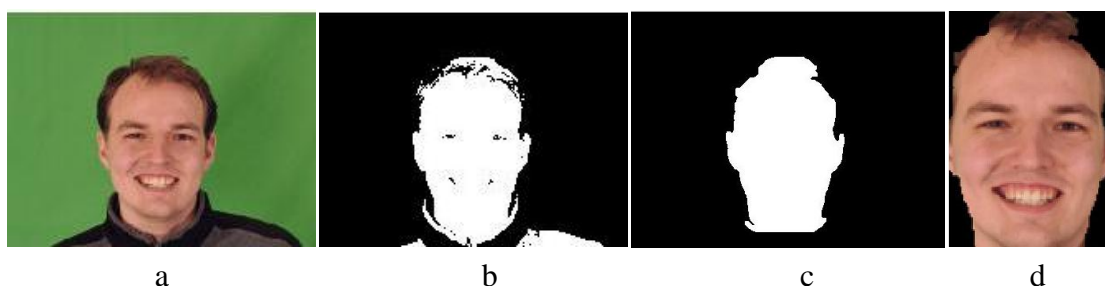
Automatický výběr oblasti obličeje je znázorněn na následujících obrázcích. Jeho správné provedení závisí na přesnosti detekce jednotlivých objektů obličeje. Na obrázku 4-4 je znázorněn zdařilý ořez obrazu z AR databáze, obrázek 4-5 uvádí méně zdařilý ořez obrazu z vlastní databáze. Ořez ze spodní strany obrazu je v pořádku, ovšem horní část obrazu zahrnuje i vlasy nad čelem a v pravé části obrazu je zachycena část ucha, nicméně obraz obsahuje celou oblast obličeje a v zásadě může být ke klasifikaci výrazu použit. Obrázek 4-6 uvádí příklad automatického ořezu, který nelze považovat za zdařilý. Výskyt takovýchto výsledků automatického ořezu je v obou databázích obrazů ojedinělý. Důvodem byla chybná detekce nosu před ořezem. Nos byl detekován na místě obočí nad pravým okem na levé straně obrazu. Z toho důvodu je část obličeje z obrazu odstraněna a na druhé straně obsahuje obraz vlasy, které by neměly být v oblasti obličeje přítomny. Detekce obličeje podle barvy navíc detekuje vlasy jako objekt s

barvou kůže. Tento obraz nevstupoval do klasifikace výrazů, protože jeho výraz není zahrnut na obrazech AR databáze, na kterých byly neuronové sítě testovány a nebylo možné tedy tento výraz rozpoznávat. Všechny automaticky ořezané obrazy vstupující do neuronové sítě pro klasifikaci výrazů obličeje obsahují v obličeji všechny významné rysy (oči, nos a ústa).

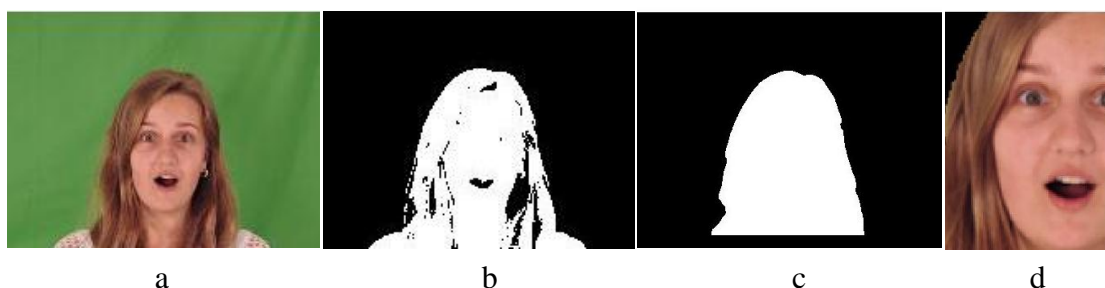
[36], [50], [63], [66]



Obrázek 4-4 Ukázka zdařilého automatického ořezu obrazu z AR databáze a) vstupní obraz, b) binární obraz po detekci bravy kůže, c) upravený binární obraz oblasti obličeje, d) obraz po automatickém ořezu



Obrázek 4-5 Ukázka méně zdařilého automatického ořezu obrazu z vlastní databáze a) vstupní obraz, b) binární obraz po detekci bravy kůže, c) upravený binární obraz oblasti obličeje, d) obraz po automatickém ořezu



Obrázek 4-6 Ukázka nezdařilého automatického ořezu obrazu z vlastní databáze a) vstupní obraz, b) binární obraz po detekci bravy kůže, c) upravený binární obraz oblasti obličeje, d) obraz po automatickém ořezu

Automatický ořez byl vizuálně porovnán s manuálním ořezem obrazů. Je patrné, že v některých případech je těžko rozhodnout, který z ořezů je přesnější. V jiných případech nebyl automatický ořez schopen vybrat oblast obličeje zcela podle předem zvolených kritérií. Příklady jsou uvedeny na obrázku 4-7.



Obrázek 4-7 Ukázka srovnání manuálního a automatického ořezu obrazů obličeje, obrázky a), c) a e) - manuální ořez, obrázky b), d), f) - automatické ořezy

Na obrázku 4-7 a) je manuální obraz osoby z AR databáze, v porovnání s automatickým ořezem na obrázku b) jsou patrné jen minimální odlišnosti. Podobně je tomu u na obrázků c) a d) z vlastní databáze. Tyto obrázky se liší detekcí hrany brady, podle které byl obraz ořezán ze spodní strany. Obrázky e) a f) jsou příkladem rozdílů mezi manuálním ořezem (e) a ořezem automatickým. V tomto případě automatický ořez vybral příliš malou oblast a hrana brady stejně horní hranice čela není v obraze přítomna.

Celý proces detekce obličeje od vstupního obrazu až po vybranou oblast obličeje je znázorněn na obrázku 4-8.



Obrázek 4-8 Vývojový diagram algoritmu pro vyhledání oblasti obličeje v barevném vstupním obrazu a následný ořez

5 ROZPOZNÁVÁNÍ VÝRAZŮ OBLIČEJE

5.1 Databáze obličejů

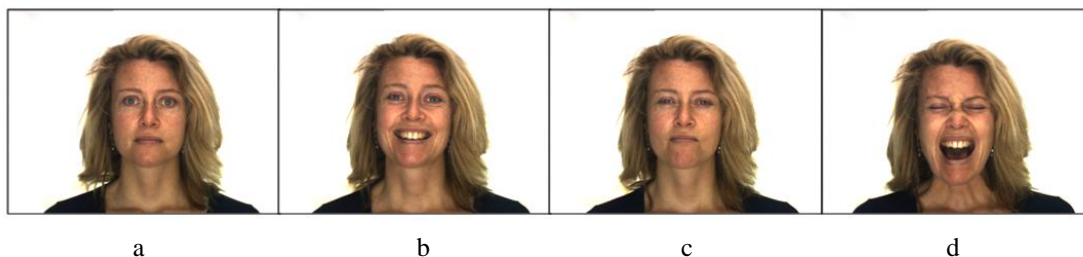
5.1.1 AR Face Database

Pro rozpoznávání výrazů obličeje byla vybrána databáze AR Face Database autorů Aleix Martinez a Robert Benavente z Computer Vision Center, The Ohio State University. Databáze obsahuje barevné obrazy 126 různých osob (70 mužů a 56 žen) při čelním pohledu (pohled směrem do fotoaparátu), maximální úklon hlavy je 15° , velikost obrazů je 768x576 pixelů s 24 bitovou barevnou hloubkou, snímky jsou uloženy ve formátu BMP. Obrazy byly pořízeny za konstantních podmínek, fotografované osoby nebyly nijak omezovány ve výběru oblečení, přítomnosti dioptrických brýlí, použití make-upu nebo zvolení účesu. [35]

Pro rozpoznávání výrazů byly využity obrazy se čtyřmi různými výrazy: neutrální výraz, úsměv, vztek a řev. Každá osoba je při každém výrazu vyfotografována dvakrát, při dvou nezávislých fotografováních s odstupem čtrnácti dnů.

Dále databáze obsahuje obrazy vybraných osob, které byly manuálně ořezány na oblast obličeje, výsledná velikost těchto obrazů je 150x105 pixelů. Jedná se o sto osob, celkem 800 obrazů pořízených během dvou nezávislých fotografování vždy se čtyřmi různými výrazy. Tyto obrazy byly použity jako dvě sady dat. Výsledky prvního focení byly použity pro trénování neuronové sítě pro rozpoznávání výrazů. Obrazy ze druhého focení byly použity pro klasifikaci výrazů bez nutného (automatického) ořezu před klasifikací.

Obrazy bez manuálního ořezu byly použity jako dvě další sady dat pro klasifikaci výrazů obličeje, každá sada dat obsahuje stejné osoby se stejnými výrazy obličejů, obrazy však pochází ze dvou nezávislých fotografování s časovým odstupem. Na obrázku 5-1 je uveden příklad obrazů databáze: je zde zobrazena jedna osoba se čtyřmi různými výrazy obličeje. [35]



Obrázek 5-1 Ukázka obrazů databáze AR s různými výrazy obličeje bez manuálního ořezu, a) neutrální výraz, b) smích, c) vztek, d) řev

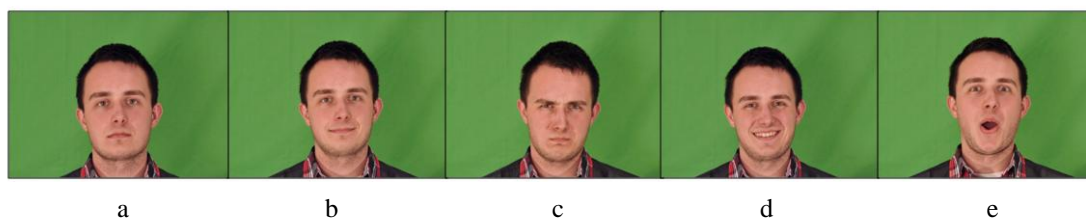
5.1.2 Vlastní databáze

Podle zadání práce byla nafocena uživatelská databáze 12 jedinců s různými výrazy obličeje. Databáze obsahuje obrazy 7 mužů a 5 žen, celkem bylo pořízeno 72 snímků. Podmínky pořizování obrazů byly voleny podobně jako podmínky obrazů z AR databáze. Osoby měly libovolné oblečení, účes, make-up či dioptrické brýle podle vlastního rozhodnutí. Foceny byly před zeleným plátnem, které bylo pro focení k dispozici, vzdálenost od fotoaparátu byla přibližně 3 metry. Osvětlení bylo zajištěno dvěma trvalými světly Fomei Digital Light 1000 W svítícími z boku a nastavenými na 60 % výkonu a jedním trvalým světlem stejného typu, které bylo doplněno rozptylným Soft Boxem IFF 1024, umístěným vedle fotoaparátu. Snímky byly pořízeny fotoaparátem Nikon D90 s objektivem 18-105mm světelnost $f/3.5-5.6$, ohnisková vzdálenost byla 50.00 mm (75 EQ), ISO 400, doba expozice 1/60s, clona 8.0.

Bylo zvoleno pět odlišných výrazů: neutrální (fotografován dvakrát; jako první výraz a jako poslední), úsměv, smích, rozzlobení, údiv. Výrazy byly voleny ve shodě s výrazy z AR databáze z výjimkou výrazu řev, který nevypadá zcela přirozeně a bylo by obtížné ho u všech osob vyfotografovat. Navíc byly přidány výrazy smích a překvapení.

Pořízené snímky byly upraveny tak, aby se přiblížily snímkům z databáze AR. Velikost snímků byla upravena na 850x580 pixelů, snímky byly uloženy ve formátu BMP s 24 bitovou barevnou hloubkou. Byla snížena saturace snímků na 73% původní hodnoty a hodnoty v červeném kanálu byly sníženy na 97%. Úpravy byly provedeny v programu Zoner Photo Studio 16, viz příloha B.

Pro klasifikaci výrazů obličeje byly použity pouze snímky s neutrálními výrazy, smích a rozzlobení, tyto tři výrazy odpovídají výrazům z AR databáze, na kterých byla neuronová síť natrénována. Malý počet vlastních snímků neumožňuje neuronovou síť na vlastních obrazech natrénovat. Všechny nafocené výrazy jsou uvedeny na obrázku 5-2.



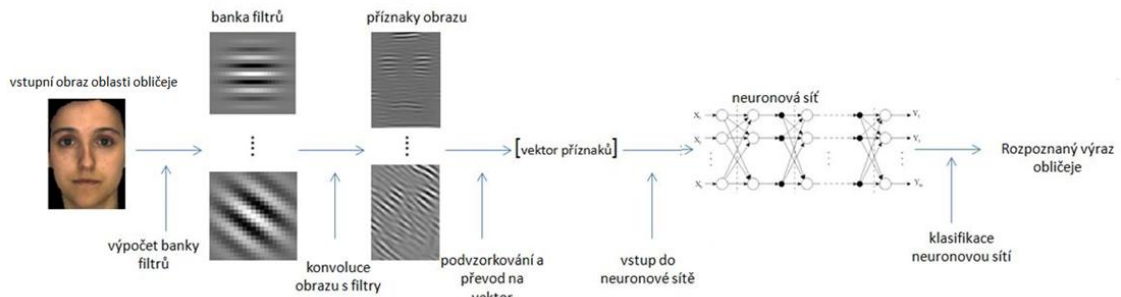
Obrázek 5-2 Ukázka obrazů vlastní databáze s různými výrazy obličeje bez manuálního ořezu, a) neutrální výraz, b) úsměv, c) vztek, d) smích (s otevřenou pusou), e) překvapení

5.2 Extrakce příznaků z obrazu

Pro rozpoznávání výrazů obličeje byl zvolen vektor příznaků získaný pomocí Gaborových filtrů. Obraz obsahující pouze oblast obličeje s poměrem stran 7:10 (poměr stran byl zvolen v souladu s obrazy z AR databáze a podle [48], [35] byl dále zpracováván pomocí 2D Gaborových filtrů za účelem získání vektoru příznaků obrazu, potřebného pro klasifikaci pomocí neuronové sítě.

Pro získání vektoru příznaků byly použity dvě banky Gaborových filtrů, každá banka obsahovala deset různých filtrů, které se lišily orientacemi a frekvencemi dvourozměrného sinusoidálního signálu. Dále byla navržena banka deseti vlastních filtrů.

Postup extrakce příznaků a následného rozpoznávání výrazů je zobrazen ve vývojovém diagramu na obrázku 5-3.



Obrázek 5-3 Vývojový diagram rozpoznávání výrazu tváře

5.2.1 První banka Gaborových filtrů

První banka deseti dvourozměrných Gaborových filtrů byla vytvořena podle následující rovnice podle [48]

$$G_{\sigma_x, \sigma_y, f, \theta}(x, y) = \exp\left(-\left[\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right]\right) \cos(2\pi f x_\theta + \varphi), \quad (5.1)$$

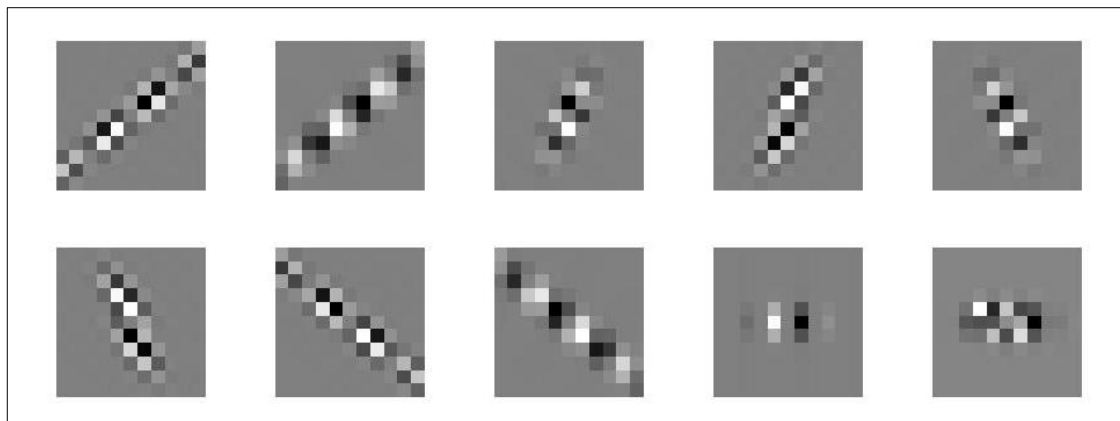
kde G je dvourozměrný Gaborův filtr s reálnými hodnotami, x a y jsou proměnné dvourozměrného prostoru, σ_x a σ_y označují směrodatné odchylky Gaussovy křivky, podél os x a y , přičemž $\sigma_x = 2$ a $\sigma_y = 1$. Proměnná f označuje centrální frekvenci sinusoidálního signálu a θ je úhel šíření sinusoidálního signálu. Hodnoty centrální frekvence a hodnoty úhlu šíření signálu byly zvoleny takto: $f \in \{0,75, 1,5\}$, $\theta \in \{\frac{\pi}{5}, \frac{2\pi}{5}, \frac{3\pi}{5}, \frac{4\pi}{5}, \frac{5\pi}{5}\}$. Fáze φ je nastavena na hodnotu $\varphi = \pi/2$. [48] Proměnné x_θ a y_θ jsou určeny následujícími rovnicemi

$$x_{\theta} = x \cos \theta + y \sin \theta , \quad (5.2)$$

$$y_{\theta} = y \cos \theta - x \sin \theta , \quad (5.3)$$

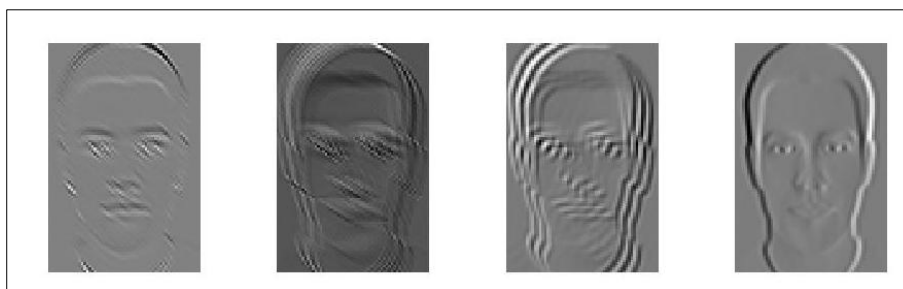
ve kterých vystupují proměnné popsané výše. [48]

Pro pět různých hodnot úhlů šíření signálu a pro dvě hodnoty centrální frekvence bylo získáno deset Gaborových filtrů, které jsou zobrazeny na obrázku 5-4.



Obrázek 5-4 Banka deseti Gaborových 2D filtrů podle rovnice 5.1

Gaborovy příznaky byly získány konvolucí šedotónových obrazů obličejů s Gaborovými filtry. Pro jeden obraz bylo konvolucí s deseti různými filtry získáno deset různých výstupních obrazů. Vybrané výstupní obrazy jsou uvedeny na obrázku 5-5. Každý z výstupních obrazů byl dále podvzorkován faktorem 10 a převeden na jednorozměrný vektor. Výsledný vektor příznaků pro jeden obraz obličeje obsahuje všechny jednorozměrné vektory získané konvolucí vstupního obrazu se všemi deseti Gaborovými filtry, celkem 700 hodnot zapsaných ve sloupcovém vektoru. [3], [48]



Obrázek 5-5 Příklad výstupních obrazů po konvoluci s Gaborovými filtry z první banky

5.2.2 Druhá banka Gaborových filtrů

Druhá banka filtrů byla realizována podle [18]. Bylo opět vytvořeno deset dvourozměrných filtrů s různou frekvencí a orientací dvourozměrného komplexního

sinusoidálního signálu, nastavení těchto proměnných bylo shodné s nastavením pro první banku filtrů. Filtry byly vytvořeny podle následující rovnice [18]

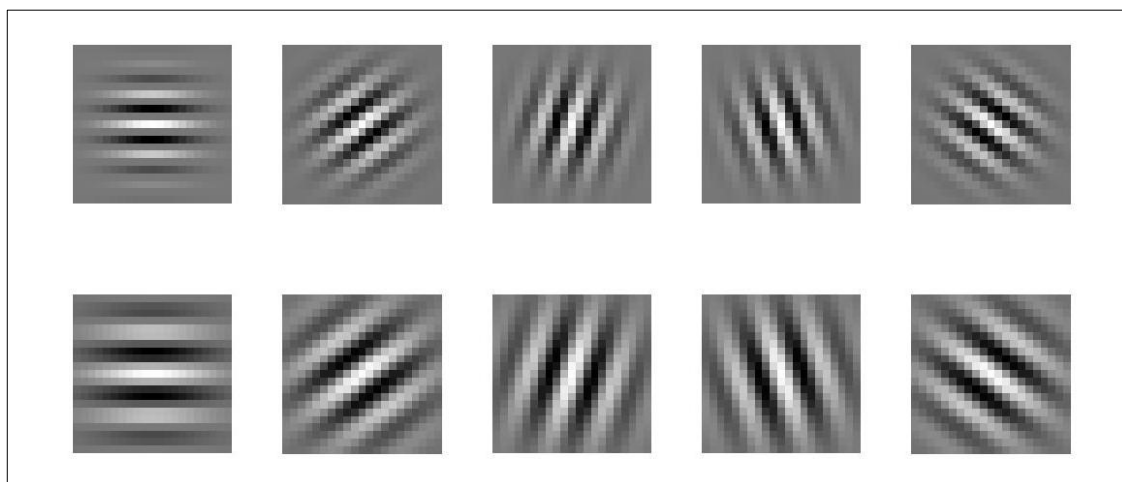
$$G_{\sigma,f,\theta}(x,y) = \frac{f^2}{\pi\gamma\eta} \exp(-[\alpha^2 x_\theta^2 + \beta^2 y_\theta^2]) \exp(j2\pi f x_\theta), \quad (5.4)$$

kde G je dvourozměrný komplexní Gaborův filtr, x a y jsou proměnné dvourozměrného prostoru, které nabývají maximálních hodnot x_{max} a y_{max} , j označuje imaginární jednotku. Proměnná f označuje střední frekvenci sinusoidálního signálu, nabývá hodnot $f \in \{0,25, 0,18\}$. Parametr α určuje strmost Gaussovy funkce podél hlavní osy, β určuje strmost Gaussovy funkce podél vedlejší osy. Dále jsou definovány proměnné $\gamma=f/\alpha$ a $\eta=f/\alpha$ k zachování poměru mezi frekvencí a špičatostí. θ představuje rotaci prostorové vlny ve směru proti hodinovým ručičkám a nabývá následujících hodnot $\theta \in \{0, \frac{\pi}{5}, \frac{2\pi}{5}, \frac{3\pi}{5}, \frac{4\pi}{5}\}$. Proměnné x_θ a y_θ jsou určeny následujícími rovnicemi

$$x_\theta = (x - \frac{x_{max}+1}{2}) \cos \theta + (y - \frac{y_{max}+1}{2}) \sin \theta, \quad (5.5)$$

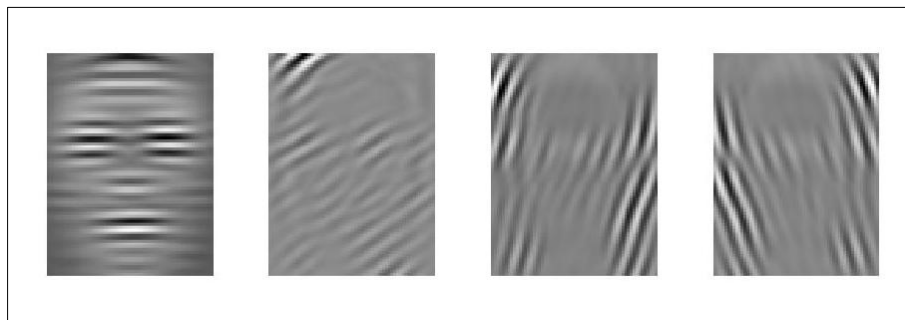
$$y_\theta = -(x - \frac{x_{max}+1}{2}) \sin \theta + (y - \frac{y_{max}+1}{2}) \cos \theta, \quad (5.6)$$

ve kterých vystupují proměnné popsané výše. Získané masky filtrů jsou zobrazeny na obrázku 5-6. [18], [48]



Obrázek 5-6 Banka deseti Gaborových 2D filtrů (reálná složka) podle rovnice 5.4

Gaborovy příznaky byly získány konvolucí šedotónových obrazů obličejů s reálnými složkami Gaborových filtrů. Výstupem připraveným pro vstup do neuronové sítě byl pro každý vstupní obraz jednorozměrný vektor obsahující výstupy všech deseti filtrů. Tento postup je podrobněji popsán v předchozích odstavcích. Následující obrázek uvádí čtyři vybrané výstupy konvoluce vstupního obrazu s Gaborovými filtry. [3], [18]



Obrázek 5-7 Příklad výstupních obrazů po konvoluci s Gaborovými filtry z druhé banky

Stejně jako v předchozím případě byl získán vektor příznaků podvzorkováním výstupních obrazů po konvoluci s jednotlivými maskami.

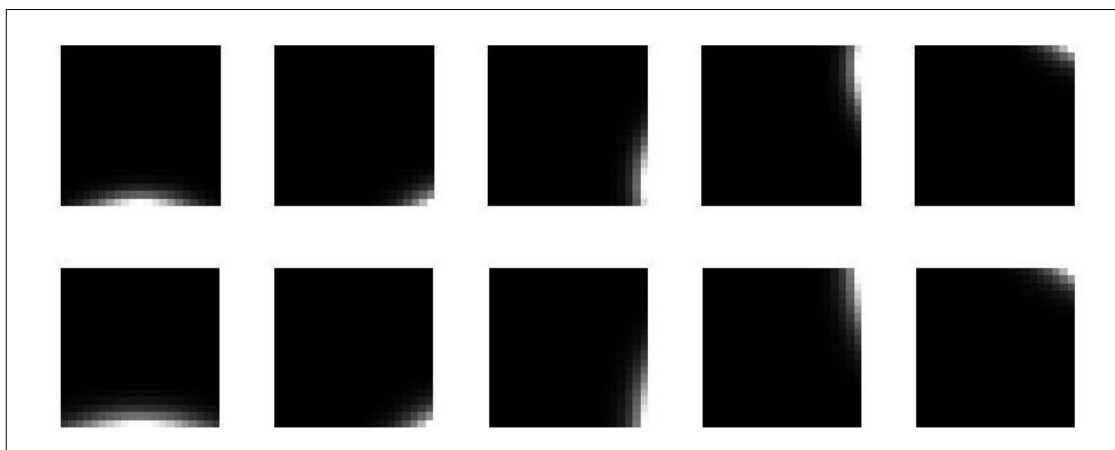
5.2.3 Vlastní návrh banky filtrů

Kromě Gaborových filtrů byla v rámci získávání příznaků z obrazů obličeje implementována další banka deseti filtrů, které lze zařadit do skupiny lineárních 2D prostorově invariantních filtrů. Masky filtrů byly vytvořeny podle následující rovnice

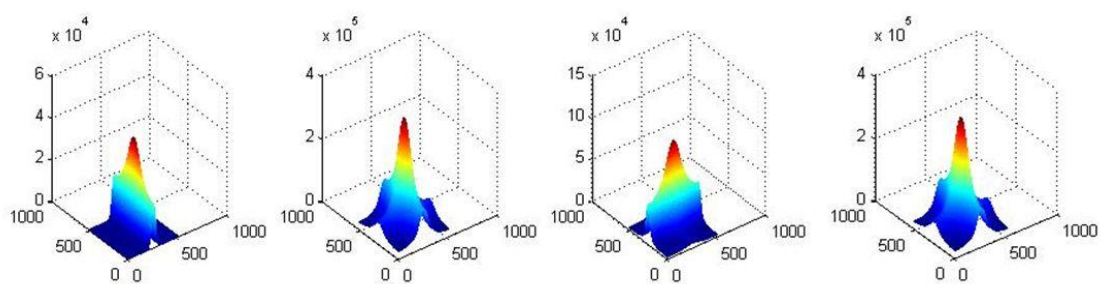
$$F_{s,\theta}(x,y) = s \exp(2\pi \sqrt{2\pi s} x_\theta - [\alpha^2 x_\theta^2 + \beta^2 y_\theta^2]) , \quad (5.7)$$

kde F je dvourozměrný filtr, x a y jsou proměnné dvourozměrného prostoru, θ představuje rotaci Gaussovy funkce ve směru proti hodinovým ručičkám, nabývá následujících hodnot $\theta \in \{0, \frac{\pi}{5}, \frac{2\pi}{5}, \frac{3\pi}{5}, \frac{4\pi}{5}\}$, proměnné x_θ a y_θ jsou určeny stejně jako u předcházejících filtrů (rovnice 6.5 a 6.6). Parametr α určuje strmost Gaussovy funkce podél hlavní osy, β určuje špičatos Gaussovy funkce podél vedlejší osy. Parametr s určuje míru rozmazání obrazu po konvoluci s bankou filtrů, jeho hodnoty byly určeny takto: $s \in \{0,005; 0,010\}$.

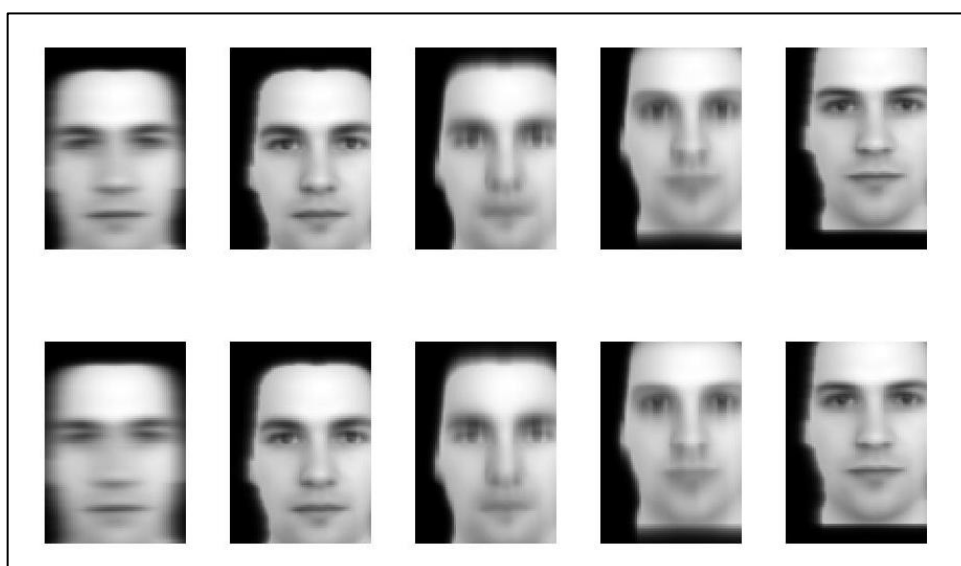
Získané masky filtrů jsou zobrazeny na obrázku 5-8., frekvenční charakteristiky vybraných filtrů jsou zobrazeny na obrázku 5-9. Z obrázku vyplývá, že se jedná o dolní propusti, ve směru podél hlavní osy mají filtry jiné vlastnosti, co se propustnosti týče, než ve směru podél vedlejší osy. Výsledky konvoluce obrazu s bankou filtrů jsou uvedeny na obrázku 5-10. V důsledku konvoluce dochází ve výsledném obraze k posunu obličeje. To je způsobeno pozicemi oblastí s nenulovými hodnotami na okrajích jednotlivých masek filtrů (viz obrázek 5-8). [18], [48]



Obrázek 5-8 Banka deseti 2D filtrů podle rovnice 5.7



Obrázek 5-9 Frekvenční charakteristiky 1., 2., 3. a 5. filtru získaných podle rovnice 5.7 po aplikaci Fourierovy transformace



Obrázek 5-10 Výsledky po aplikaci vlastní banky filtrů

5.3 Neuronová síť

K naučení neuronové sítě byla použita sada 400 obrazů z AR databáze, které jsou manuálně ořezány na oblast obličeje. Do neuronové sítě vstupovaly vektory příznaků trénovacích obrazů, získané konvolucí obrazů s vybranou bankou filtrů. Výstupy neuronové sítě byly porovnány s očekávanými výstupy. Učení neuronové sítě probíhalo pomocí algoritmu Error Back Propagation, který je blíže popsán v kapitole 3.2.2.

Vstupními parametry neuronové sítě jsou počet skrytých vrstev neuronů a počet neuronů v jednotlivých skrytých vrstvách. Počet neuronů vstupní vrstvy je dán počtem hodnot ve vektoru příznaků každého vstupního obrazu (v tomto případě 700 hodnot), počet neuronů ve výstupní vrstvě je roven počtu tříd, do kterých mohou být výrazy obličejů na obrazech klasifikovány (v tomto případě 4). Byly testovány různé počty skrytých vrstev neuronů a různý počet neuronů ve skrytých vrstvách. Z důvodů paměťových omezení použitého počítače nebylo možné vytvořit neuronovou síť s více než 100 neurony v jedné skryté vrstvě ani více než 2 skryté vrstvy. Výsledky testování pro různé vektory příznaků (získané pomocí různých bank filtrů) jsou uvedeny v tabulkách 5-1 až 5-3. Tabulky uvádějí nejlepší výsledky dosažené během 10 opakování učení sítě s danými počty skrytých vrstev a počty neuronů v těchto vrstvách. Výsledky během deseti opakování se mohou lišit až o 20%, značný vliv má v procesu učení náhoda. Pro dostatečné posouzení by bylo potřeba proces učení pro určité nastavení sítě opakovat řádově stokrát nebo více. To nebylo z časových důvodů možné, proces 100 násobného učení jedné sítě by trval cca 12 hodin, pro 15 různých nastavení sítě a 3 různé vektory příznaků by testování trvalo zhruba 540 hodin výpočetního času, což je 22 dní.

Sada trénovacích obrazů (označena jako Sada1) byla rozdělena na tři skupiny obrazů, první skupina se 70% vstupních obrazů byla použita pro učení sítě. Druhá a třetí skupina obsahovaly každá 15% vstupních obrazů. Druhá skupina byla použita pro validaci naučení sítě, třetí skupina byla použita pro testování sítě. Rozřazení vstupních obrazů do jednotlivých skupin bylo pro každou z vytvořených sítí shodné, šlo o snahu snížit vliv náhody na rozřazení obrazů a tím i na výsledky správnosti klasifikace sítě. Úspěšnost naučení sítě lze hodnotit podle dosažené úspěšnosti klasifikace vstupních obrazů, nebo podle parametru *performance*, který udává výkonnost sítě a je také výstupem procesu trénování neuronové sítě. Parametr je počítán funkcí *perform* se znalostí očekávaných a skutečných výstupů sítě. Hodnoty *performance* se pohybují v intervalu $<0,1>$, hodnoty blízké nule znamenají dobře naučenou síť, která správně klasifikuje velké procento vstupních obrazů.

Tabulka 5-1, 5-2 a 5-3 uvádí počty neuronů ve skrytých vrstvách sítě, *performance*,

procenta správně klasifikovaných výrazů ze skupiny určené na učení (70% obrazů) a ze skupiny určené na testování (15% obrazů). [10]

Tabulka 5-1 Výsledky testování počtů skrytých vrstev sítě a počtů neuronů ve skrytých vrstvách pro neuronové sítě učené na vektorech příznaků získaných pomocí 1. banky filtrů

Počet neuronů první skryté vrstvy	Počet neuronů druhé skryté vrstvy	Performance	Úspěšnost trénování [%]	Úspěšnost testování [%]
50	10	0,0430	88,6	95,0
50	20	0,0421	90,4	88,3
50	30	0,0414	90,7	83,3
50	40	0,0462	88,2	91,7
75	10	0,0419	90,7	91,7
75	20	0,0521	89,0	73,3
75	30	0,0523	88,3	71,7
75	40	0,0504	89,5	68,3
100	10	0,0425	91,1	86,7
100	20	0,0322	92,1	95,0
100	30	0,0422	88,9	86,7
100	40	0,0421	91,8	86,7
50	-	0,0446	90,4	86,7
40	-	0,0370	90,4	91,7
30	-	0,0451	89,6	83,3
20	-	0,0453	90,7	83,3
10	-	0,0430	89,6	96,7

Tabulka 5-2 Výsledky testování počtů skrytých vrstev sítě a počtů neuronů ve skrytých vrstvách pro neuronové sítě učené na vektorech příznaků získaných pomocí 2. banky filtrů

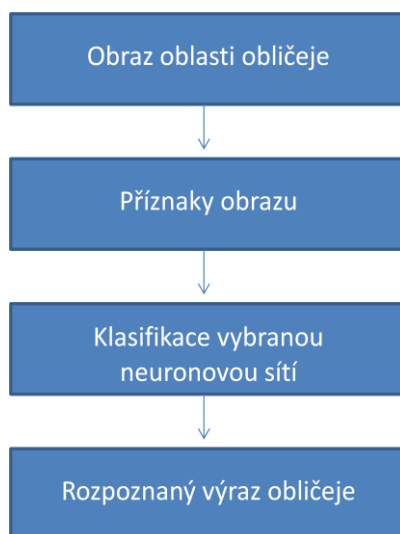
Počet neuronů první skryté vrstvy	Počet neuronů druhé skryté vrstvy	Performance	Úspěšnost trénování [%]	Úspěšnost testování [%]
50	10	0,0544	88,2	85,0
50	20	0,0460	90,7	83,3
50	30	0,0538	86,4	91,7
50	40	0,0527	86,8	88,3
75	10	0,0550	86,1	86,7
75	20	0,0649	87,9	83,0
75	30	0,0713	83,9	78,3
75	40	0,0631	85,7	86,7
100	10	0,0667	82,9	81,7
100	20	0,0742	83,2	83,3
100	30	0,0510	86,4	86,7
100	40	0,0537	87,1	90,0
50	-	0,0538	87,5	86,7
40	-	0,0455	90,7	83,3
30	-	0,0511	88,6	83,3
20	-	0,0458	88,9	86,7
10	-	0,0498	87,1	90,0

Tabulka 5-3 Výsledky testování počtů skrytých vrstev sítě a počtů neuronů ve skrytých vrstvách pro neuronové sítě učené na vektorech příznaků získaných pomocí 3. banky filtrů

Počet neuronů první skryté vrstvy	Počet neuronů druhé skryté vrstvy	Performance	Úspěšnost trénování [%]	Úspěšnost testování [%]
50	10	0,0448	90,7	88,3
50	20	0,0358	93,9	88,3
50	30	0,0455	91,1	85,0
50	40	0,0503	87,5	88,3
75	10	0,0436	89,3	91,7
75	20	0,0455	88,2	88,3
75	30	0,0433	91,8	91,7
57	40	0,0430	88,8	83,3
100	10	0,0476	89,3	91,7
100	20	0,0416	90,7	91,7
100	30	0,0393	91,4	90,0
100	40	0,0486	88,9	91,7
50	-	0,0491	88,6	93,3
40	-	0,0512	89,6	86,7
30		0,0498	89,3	85,0
20	-	0,0442	90,7	90,0
10	-	0,0435	88,6	83,3

5.4 Výsledky rozpoznávání výrazu tváře

Pro výrazy obličejů na vstupních obrazech, které měly být klasifikovány, následovalo po výběru oblasti obličeje a po výpočtu příznaků obrazu pomocí zvolené banky filtrů rozpoznávání výrazu zvolenou neuronovou sítí. Vektor příznaků obrazu vstupoval do neuronové sítě a výstupem bylo zařazení výrazu obličeje v obraze do některé ze čtyř skupin klasifikovaných výrazů. V případě, že byl znám očekávaný výsledek rozpoznání výrazu, byl očekávaný výsledek porovnán s dosaženým výsledkem a součástí výstupu bylo sdělení, zda se očekávaný a získaný výsledek shoduje, nebo ne. Uvedený postup je zobrazen na obrázku 5-11.



Obrázek 5-11 Vývojový diagram procesu rozpoznávání výrazu obličeje v obraze

Tato kapitola uvádí výsledky klasifikace výrazů obličejů v testovaných obrazech pomocí neuronových sítí. Parametry nejlépe naučených sítí pro každou ze tří variant banky filtrů byly uloženy a takto naučená síť mohla být použita pro klasifikaci obličejů na neznámých obrazech. Dostupné obrazy byly rozděleny do několika skupin. Obrazy z AR databáze tvoří 4 skupiny obrazů. První skupina (Sada1) byla použita k naučení neuronové sítě. Druhá skupina (Sada2) obsahovala ručně ořezané obrazy pouze s oblastí obličeje, ve skupině bylo celkem 400 obrazů, byly to obrazy stejných osob jako ve skupině 1, jednalo se o druhou fázi focení (odstup od prvního focení minimálně 14 dní). Sada 3 a 4 obsahovaly fotografie 135 osob včetně pozadí, celkem tedy 540 obrazů. Tyto obrazy byly nejprve zpracovány automatickým ořezem, poté vstupovaly do neuronové sítě. Poslední dvě skupiny obrazů (Sada 5 a 6) obsahují vlastní databázi snímků, která obsahuje 48 obrazů 12 různých osob před pozadím. U sady 5 byla oblast obličeje vybrána ručně, sada 6 byla zpracována automatickým ořezem podle algoritmu

popsaného výše a poté byly obě sady klasifikovány neuronovou sítí. Výsledky klasifikace obrazů byly porovnány s předpokládanými výsledky, úspěšnost klasifikace pro vybrané neuronové sítě pro různé skupiny obrazů jsou uvedeny v tabulkách 5-4 až 5-6.

Mezi sítěmi naučenými na vektorech příznaků vypočítaných konvolucí s první bankou filtrů dosahovala nejlepších výsledků síť se dvěma skrytými vrstvami se 100 neurony v první skryté vrstvě a 20 neurony ve druhé vrstvě viz Tabulka 5-1. Výsledky klasifikace výrazů jsou uvedeny v tabulce 5-4.

Tabulka 5-4 Úspěšnost klasifikace pro různé sady obrazů klasifikované sítí se 2 skrytými vrstvami (100, 20), příznaky získány pomocí první banky filtrů

Sada obrazů	% správně zařazených výrazů
Sada 2 AR	75,0
Sada 3 AR	33,0
Sada 4 AR	29,7
Sada 5 Vlastní	38,3
Sada 6 Vlastní	29,2

Mezi sítěmi naučenými na vektorech příznaků vypočítaných konvolucí s druhou bankou filtrů se tři nejlépe natrénované sítě lišily v parametru performance v rozsahu 7 desetitisícin, viz tabulka 5.2. Z důvodu nejlepšího výsledku v testování sítě a nejmenšího rozdílu výsledků mezi trénováním a testováním sítě byla vybrána síť s jednou skrytou vrstvou obsahující 20 neuronů. Výsledky klasifikace výrazů jsou uvedeny v tabulce 5-5.

Tabulka 5-5 Úspěšnost klasifikace pro různé sady obrazů klasifikované sítí s jednou skrytou vrstvou se 20 neurony, příznaky získány pomocí druhé banky filtrů

Sada obrazů	% správně zařazených výrazů
Sada 2 AR	69,5
Sada 3 AR	29,1
Sada 4 AR	35,6
Sada 5 Vlastní	25,0
Sada 6 Vlastní	20,8

Mezi sítěmi naučenými na vektorech příznaků vypočítaných konvolucí s třetí bankou filtrů dosahovala nejlepšího výsledku síť se dvěma skrytými vrstvami se 100 neurony v první skryté vrstvě a 30 neurony ve druhé vrstvě, viz Tabulka 5-3. Výsledky klasifikace výrazů jsou uvedeny v tabulce 5-6.

Tabulka 5-6 Úspěšnost klasifikace pro různé sady obrazů klasifikované sítí se 2 skrytými vrstvami (100, 30), příznaky získány pomocí třetí banky filtrů

Sada obrazů	% správně zařazených výrazů
Sada 2 AR	81,5
Sada 3 AR	59,1
Sada 4 AR	54,1
Sada 5 Vlastní	38,3
Sada 6 Vlastní	37,5

Nejlepší výsledky klasifikace výrazů dává neuronová síť naučená na výstupech vlastní banky filtrů. Tato síť dosahuje nejlepších výsledků v porovnání s ostatními sítěmi pro všechny sady testovaných obrazů, pro sadu obrazů číslo 5 dosahuje tato síť stejných výsledků jako síť naučená na výstupech filtrů banky 1.

Pro učící množinu dat vykazovala síť naučená na výstupech vlastní banky filtrů přesnost 91,5% (pro všechny obrazy - na trénování, testování a validaci společně). Největší nepřesnosti klasifikace výrazů na trénovací množině obrazů byly způsobeny záměnou mezi výrazy 1 a 3, tedy mezi neutrálním výrazem a vztekem. Tato skutečnost je logická, neboť výrazy skupiny 1 a 3 jsou si blízké, zatímco výrazy smích a křik se výrazněji odlišují. Obrazy s výrazy 2 a 4 (smích a křik) byly správně rozpoznány v 99% případů. Nejmenší přesnost rozpoznání výrazu vykazovala skupina výrazu 3, viz příloha A.1.

Nejlepší výsledky klasifikace v rámci testování jedné neuronové sítě dává sada obrázků číslo 2 - okolo 80%. Tento výsledek potvrdil předpoklad, že sada 2 bude klasifikována nejpresněji. Obrazy této sady jsou na rozdíl od jiných sad obrazů trénovací sadě velmi podobné, protože se jednalo o databázi stejných obličejů jako trénovací množina, jen byly snímány s časovým odstupem. Obrazy byly upraveny ručním ořezem oblasti obličeje. Tím byly eliminovány nepřesnosti vznikající u jiných sad obrazů chybným nebo nepřesným ořezem obrazů. Stejně jako v trénovací množině dat i zde byly nejčastěji zaměněny výrazy 1 a 3 (neutrální výraz a vztek), výraz 3 vykazoval nejmenší přesnost klasifikace (71%). Nejpresnější výsledky klasifikace byly dosaženy pro výraz 4 - křik, viz příloha A.1.

Třetí a čtvrtá sada dat naopak vykazovala nejpřesnější výsledky pro výraz 3 - vztek. Celkově klasifikace dosáhla přesnosti 59,1% pro třetí sadu obrazů a 54,1% pro čtvrtou sadu. Falešná detekce výrazu 3 nastala velmi často, zatímco falešná detekce výrazu 2 velmi zřídka. Tomu odpovídá, že výraz 2 byl klasifikován s největší přesností a výraz 3 s nejmenší přesností.

U skupiny dat 5 a 6 se výraz 4 nevyskytoval, přesto byly některé obrazy do této skupiny výrazů chybně zařazeny (nejčastěji to byly obrazy s neutrálním výrazem). Největší přesnost detekce vykazovaly obrazy s výrazem 3 - vztek. Neutrální výraz byl nejčastěji chybně klasifikován. Celkově se přesnost klasifikace blížila hodnotě 40%, viz příloha A.1.

Obecně platí, že nejméně falešných detekcí vzniklo pro výraz 2 - smích. Pro výraz 4 - křik vzniklo také relativně málo falešných detekcí. Falešné detekce nejčastěji výraz obličeje přiřadily k výrazu 2 - smích. Nejčastěji zaměňovány byly výrazy 1 a 3 - neutrální výraz a vztek.

Trénovací sada obrazů dávala nejlepší výsledky klasifikace, což se z povahy metody neuronové sítě předpokládalo. U obrazů sady 3 a 4 je menší přesnost detekce způsobena vlivem automatického ořezu obrazů a jeho nepřesností. Naopak v sadě dat 5 a 6 je rozdíl výsledků mezi obrazy ořezanými ručně a automaticky velmi malý, viz příloha A.1.

6 ZÁVĚR

V teoretické části této práce byly popsány základní charakteristiky barevných obrazů, metody detekce obličejů v obrazech případně ve videu a metody rozpoznávání výrazů obličejů v obrazech.

V praktické části byl popsán postup detekce obličeje, byla zvolena metoda založená na detekci barvy kůže. Detekce byla prováděna ve třech barevných prostorech: RGB, HSV, YCbCr. Důvody pro zvolení této metody byly její jednoduchá implementace, výpočetní rychlost a fakt, že tato metoda nevyžaduje učení. Detekce obličeje byla doplněna o hledání základních rysů obličeje (oči, nos a ústa). Pomocí těchto rysů docházelo k rozměření oblasti obličeje a následný automatický ořez oblasti obličeje byl proveden podle hranic obličeje vypočtených z pozic základních rysů. Automatický ořez proběhl bez většího selhání na všech obrazech obou databází. V některých případech však manuálně ořezané obrazy lépe vystihují oblast obličeje.

Pro rozpoznávání výrazů byla vybrána metoda založená na analýze příznaků obrazu. Vektor příznaků byl získán filtrací obrazu bankou filtrů. Práce obsahuje dvě banky Gaborových filtrů a jednu banku s vlastními navrženými filtry. Klasifikace výrazů obličeje byla provedena pomocí neuronové sítě, do které vstupuje vektor příznaků. Výraz obličeje byl přiřazen k jednomu ze čtyř předem určených výrazů. Neuronová síť byla naučena na obrazech z AR databáze. Součástí práce bylo testování parametrů neuronové sítě.

Práce uvádí dosažené výsledky rozpoznávání výrazů obličeje na obrazech z AR databáze a na obrazech z vlastní databáze, která byla realizovaná v rámci diplomové práce.

Síť, která dosahuje nejlepších výsledků v oblasti učení, byla síť trénovaná na výstupech první banky Gaborových filtrů, pro trénovací množinu dat dosahuje přesnosti 93%. Ovšem nejlepších výsledků rozpoznávání výrazu na neznámých vstupních dosahuje síť trénovaná na bance vlastních filtrů, jejíž výsledky ve fázi učení dosahují v porovnání s ostatními sítěmi horších výsledků (91,5%). Tato síť dosahuje na manuálně ořezaných obrazech AR databáze dosahuje výsledků 81,5%, zatímco síť trénovaná na výstupech první banky Gaborových filtrů dosahuje přesnosti 75%. Na přesnost rozpoznávání výrazů má vliv způsob výběru oblasti obličeje. Obrazy zpracované automatickým ořezem vykazují horší výsledky než obrazy ořezané manuálně.

Na základě dosažených výsledků lze říct, že metoda rozpoznávání výrazu obličeje realizovaná klasifikací vektoru příznaků pomocí neuronové sítě funguje. Pro vektor

příznaků získaný pomocí banky Gaborových filtrů nedává metoda velmi dobré výsledky. Klasifikace vektoru příznaků získaného pomocí banky vlastních filtrů dosahuje větší přesnosti. Ovšem pro případné využití metody pro klasifikaci neznámých snímků (tak jako to odpovídá sadám obrazů 5 a 6), na kterých není algoritmus trénován, nejsou obdržené výsledky dostatečně přesné.

Využití některé z metod redukce objemu dat (např. PCA) pro vektor příznaků by mohlo vézt ke zlepšení výsledků rozpoznávání výrazů tváře. Také použitím jiného algoritmu pro samotnou klasifikaci než neuronových sítí by mohlo být přínosné z pohledu správnosti klasifikace výrazu tváře.

LITERATURA

- [1] BAGHERIAN, E., Wirza, R., RAHMAT, O. K. Facial feature extraction for face recognition: a review. *2008 International Symposium on Information Technology* [online]. IEEE, 2008, 2014(3), 1-9 [cit. 2016-05-08]. DOI: 10.1109/ITSIM.2008.4631649. ISBN 978-1-4244-2327-9. ISSN ISSN 2278-6856. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4631649>
- [2] BAKSHI, U., SINGHAL, R. A SURVEY ON FACE DETECTION METHODS AND FEATURE EXTRACTION TECHNIQUES OF FACE RECOGNITION. *International Journal of Emerging Trends & Technology in Computer Science* [online]. 2014, 2014(3) [cit. 2016-05-08]. ISSN ISSN 2278-6856. Dostupné z: <http://www.mihantarjomeh.com/wp-content/uploads/2015/01/A-SURVEY-ON-FACE-DETECTION.pdf>
- [3] BARBU, T. Gabor Filter-Based Face Recognition Technique. *The Publishing House of the Romanian Academy* [online]. 2008, 2010(3), s. 227-283 [cit. 2016-05-08]. ISBN 978-1-4244-2327-9. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4631649>
- [4] BARTONČÍK, M. *Rozpoznávání výrazu tváře u neznámých osob*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2011. 68s. Vedoucí diplomové práce byl Ing. Karel Horák, Ph.D.
- [5] BAŘINA, D. Gabor Wavelets in Image Processing, Brno. Vysoké učení technické v Brně, Fakulta informačních technologií, 2009, 5 s. Vedoucí práce: Ing. Pavel Zemčík
- [6] BETTADAPURA, Vinay. *Face Expression Recognition and Analysis:: The State of the Art*. [online]. 2003, : 27 [cit. 2016-01-03].
- [7] BHUIYAN, A., LIU, Ch. H. On face recognition using gabor filters. *World academy of science, engineering and technology*, 2007, 28: s. 51-56.
- [8] CAI, Zhihua. Computational intelligence and intelligent systems: 4th International Symposium, ISICA 2009, Huangshi, China, October 23-25, 2009 : proceedings. New York: Springer, 2009, xiii, 484 p. Communications in computer and information science, 51. ISBN 36-420-4961-3.
- [9] Color Representations. *Colgate University: Computer Science* [online]. 2012 [cit. 2016-01-03]. Dostupné z: <http://cs.colgate.edu/~chris/cs122web/Unit%201%20P.E/Color%20Representations.html>
- [10] DAVALO, E., NAIM, P. *Neural networks*. Houndmills: Macmillan Education, 1991. ISBN 03-335-4996-1.
- [11] DING, L., MARTINEZ, A. M. Features versus context: An approach for precise and detailed detection and delineation of faces and facial features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(11), 2010. s. 2022-2038.
- [12] *Face Detection* [online]. 1999, 2016 [cit. 2016-01-03]. Dostupné z: <https://facedetection.com/techniques>
- [13] FASEL, B. a J. LUETTIN. Automatic Facial Expression Analysis: A Survey. *Pattern*

- Recognition* [online]. 2002, 2003(36), s. 259-275 [cit. 2016-05-08]. ISBN 978-1-4244-2327-9. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4631649>.
- [14] FRANC, Vojtěch. *Support vector machines (SVM): Algoritmy podpůrných vektorů* [online]. Brno, 2005, : 3 [cit. 2016-01-03].
- [15] *Grafický formát BMP* [online]. Praha, 2006, 2006-11-19 [cit. 2016-01-03]. Dostupné z: <http://www.root.cz/clanky/graficky-format-bmp-pouzivany-a-pritom-neoblíbeny/>
- [16] GROSS, R. MATTHEWS, I. COHN, J. KANADE, T., BAKE, S.r. *Multi-pie. Image and Vision Computing*, 2010.
- [17] GUPTA, Varsha SHARMA, Dipesh. *A Study of Various Face Detection Methods* [online]. 2004, : 4 [cit. 2016-01-03]. ISSN 2278-1021.
- [18] Haghighat, M., Zonouz, S., Abdel-Mottaleb, M. "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905-7916, 2015.
- [19] Harabiš, V. *Vizualizace medicínských dat*, Brno, 2015: 31 [cit. 2016-01-03].
- [20] HEMALATHA, G., SUMATHI, C. P. A Study of Techniques for Facial Detection and Expression Classification. *International Journal of Computer Science and Engineering Survey*, 2014, 5(2): s. 16-27.
- [21] HJELM, Erik LOW, Boon K. Face Detection: A Survey. *Computer Vision and Image Understanding* [online]. 2001-04-17, 2001(83): 39.
- [22] HÖLL, K. *Aplikace metod detekce a rozpoznání obličeje*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2014. 73 s. Vedoucí diplomové práce Ing. Petr Petyovský
- [23] HUANG, De-Shuang (ed.). *Advanced intelligent computing theories and applications: with aspects of theoretical and methodological issues : 4th International Conference on Intelligent Computing, ICIC 2008 : Shanghai, China, September 15-18, 2008 : proceedings*. Berlin: Springer, 2008, xxvii, 1273 s. Lecture notes in computer science, 5226. ISBN 978-3-540- 87440-9
- [24] CHANDRAPPA, D. N. RAVISHANKAR, M. RAMESHBABU, D. R. *Face Detection in Color Images using Skin Color Model Algorithm based on Skin Color Information* [online]. India, 2007, : 5 [cit. 2016-01-03].
- [25] [CHAPMAN, Nigel P, CHAPMAN Jenny. *Digital multimedia*. 2nd ed. Chichester: Wiley, 2004, xix, 679 p. ISBN 04-708-5890-7.
- [26] INSEONG, Kim SHIM, Joon Hyung YANG, Kinkyu. *Face detection* [online]. China, 2009, : 32 [cit. 2016-01-03].
- [27] Introduction to Colour Spaces. *Colour Management from Image to Print* [online]. 2015, 2015-12-06 [cit. 2016-01-03]. Dostupné z: http://www.colourphil.co.uk/xyz_colour_space.shtml
- [28] JÄHNE, Bernd. *Digitale bildverarbeitung*. 5. Aufl. Berlin: Spring-Verlag, 2002, 618 s. ISBN 35-404-1260-3
- [29] KELBY, Scott. *Digitální fotografie ve Photoshopu CS*. Vyd. 1. Brno: Computer Press, 2004, x, 372 s. ISBN 80-251-0482-6.

- [30] KOVAČ, Jure PEER, Peter SOLINA, Franc. *Face Detection, Pose Estimation, and Landmark Localization in the Wild*[online]. Ljubljana, Slovenia, 2003, : 8 [cit. 2016-01-03].
- [31] KŘÍKLAVÁ, Linda VALA, Ondřej. *FORMÁTY PRO RASTROVOU GRAFIKU* [online]. Praha, 2011, 2011-03-17 [cit. 2016-01-03]. Diplomová práce. ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE FAKULTA STAVEBNÍ. Vedoucí práce Miroslav Ott.
- [32] LEE, Hua WADE, Glen. *Imaging technology*. Vyd. 1. New York: Institute of Electrical and Electronics Engineers, 1986, viii, 535 p. ISBN 08-794-2199-1.
- [33] L. C. De Silva, K. Aizawa, and M. Hatori, Detection and tracking of facial features by using a facial feature model and deformable circular template, *IEICE Trans. Inform. Systems* E78–D(9), 1995, 1195–1207.
- [34] LU, Yongzhong ZHOU, Jingli YU, Shengsheng. *A SURVEY OF FACE DETECTION, EXTRACTION AND RECOGNITION*[online]. China, 2003 [cit. 2016-01-03].
- [35] MARTINEZ, A.M., BENAVENTE, R. *The AR Face Database*. CVC Technical Report #24, June 1998
- [36] *MathWorks: Matlab Academy* [online]. 1994, 2016 [cit. 2016-05-05]. Dostupné z: <http://www.mathworks.com/>
- [37] MATTHEWS, Iain BARKER, Simon. Active appearance models revisited. *IJCV*, 60(2): s. 135–164, 2004.
- [38] Morphological Image Processing. *The University of Auckland* [online]. New Zeland [cit. 2016-05-08]. Dostupné z: <https://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/ImageProcessing-html/topic4.htm>
- [39] [PETLAND, A., MOGHADDAM, B., STARNER, T. “View-Based and modular eigenspaces for face recognition,” *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pp. 84-91, 1994]
- [40] PETROU, Maria PETROU, Costas. *Image processing: the fundamentals*. 2nd ed. Chichester: Wiley, 2010, xxiii, 794 s. ISBN 978-0-470-74586-1.
- [41] MALACH, Tobiáš BAMBUCH, Petr MALACH, Jindřich. *Detekce obličejů v obraze s využitím prostředí MATLAB* [online]. Brno, 2012, výzkumná zpráva EBIS, spol. s r.o.: 6 [cit. 2016-01-03].
- [42] MARCHAND-MAILLET, S., SHARAIHA Y. M. *Binary digital image processing: a discrete approach*. San Diego: Academic Press, 2000. ISBN 01-247-0505-7.
- [43] MAURIUS, D., PENNATHUR, S., & ROSE, K. Face detection using color thresholding and eigenimage template matching. *Digital Image Processing project*. 2003,. s. 1-17
- [44] NILSSON, V., UBILLUS, A., TALOVIC, M. *Face Recognition in Color Images using Matlab TNM034*, [online]. 2012 [cit. 2016-05-08]. ISBN 978-1-4244-2327-9. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4631649>
- [45] POSPÍŠIL, L. *Detekce obličejů v obraze*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky, 2014. 44 s. Vedoucí práce Ing. Martin Mézl
- [46] RAHMAN, Abdul WEI, Kit Chong SEE, John. *RGB-H-CbCr Skin Colour Model for Human Face Detection* [online]. 2001, : 6 [cit. 2016-01-03].

- [47] RAWLINSON, Tim BHALERAO, Abhir WANG , Li. *Principles and Methods for Face Recognition and Face Modelling*[online]. UK, 2009, 2014(3): 29 [cit. 2016-01-03]. ISSN 2278-1021.
- [48] ROUHI, R., AMIRI, M., IRANNEJAD, B. A Review on Feature Extraction Techniques in Face Recognition. *Signal & Image Processing: An International Journal (SIPIJ)*. 2012, 3(6), s. 1-14. DOI: 10.5121/sipij.2012.3601. ISSN 22293922. Dostupné z: <http://www.airconline.com/sipij/V3N6/3612sipij01.pdf>
- [49] RUE, H. and HELD, L. Gaussian Markov random fields: theory and applications, volume 104. Chapman & Hall, 2005.
- [50] SAYANTANU, Paul SAYANTANU, Thakur ANKUR, Mondal. *Face Detection Using Skin Tone Segmentation* [online]. India, 2005, : 8 [cit. 2016-01-03].
- [51] *School of computer science and communication* [online]. 2011 [cit. 2011-10-16]. Vahid Kazemi. Dostupné z WWW: <<http://www.csc.kth.se/~vahidk/>>.
- [52] STAWSKA, Z., MILCZARSKI, P. Algorithms and Methods Used in Skin and Face detection Suitable for Mobile Applications. *Information Systems in Management*. Poland, 2012, 2013(2), s. 227-238.
- [53] SZNITMAN, Raphael JEDYNAK, Bruno. Active Testing for Face Detection and Localization. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, [online]. 2008, 2010(20): 7 [cit. 2016-01-03].
- [54] TAUBMAN, David S, MARCELLIN, Michael W. *JPEG2000: Image Compression Fundamentals, standards and practice*. United States of America, 2002. ISBN 07-923-7519-X
- [55] The YIQ Color Model. *Computer Graphics* [online]. 2003, 2003-07-23 [cit. 2016-01-03]. Dostupné z: <http://www.mat.univie.ac.at/~kriegl/Skripten/CG/node14.html>
- [56] TROJANOVÁ, Jana. *Automatické rozpoznání výrazu tváře s libovolným natočením v prostoru* [online]. Plzeň [cit. 2016-01-03]. Dizertační práce. Západočeská univerzita v Plzni Fakulta aplikovaných věd. Vedoucí práce Doc. Ing. Miloš Železný, Ph.D.
- [57] VIOLA, Paul; JONES, Michael. Robust Real-Time Face Detection. In *International Journal of Computer Vision*. The Netherlands: Kluwer Academic Publishers, 2003. s. 137-154.
- [58] VLACH, Jan PŘENOSIL, Jiří. *Lokalizace obličeje v obraze s komplexním pozadím*. Elektrotechnika [online]. 2010, 2007-04-11, 2007(12): 12 [cit. 2015-12-10]. ISSN 1213-1539.
- [59] WISKOTT, L. FELLOUS, J.-M. KUIGER, N. von der MALSBERG, C.. Face recognition by elastic bunch graph matching. *IEEE TPAMI*, Jul 1997.
- [60] WU, T., FU, S. YANG, G. *Survey of the Facial Expression Recognition Research* [online]. China [cit. 2016-01-03].
- [61] WU, T., Fu, S., YANG, G. *Survey of the Facial Expression Recognition Research* [online]. *Advances in Brain Inspired Cognitive Systems*. Springer Berlin Heidelberg, 2012, s. 392-402, [cit. 2016-05-08]. ISBN 978-1-4244-2327-9. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4631649>
- [62] YANG, M. H., KRIEGMAN D. J., AHUJA, N. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002, vol. 24, no. 1, p. 34-58.
- [63] ZAFEIRIOU, S., ZHANG, C., ZHANG, Z. A Survey on Face Detection in the wild: past,

- present and future, *Computer Vision and Image Understanding* (2015), Dostupné z: <http://dx.doi.org/10.1016/j.cviu.2015.03.015>
- [64] ZHANG, Ch ZHANG, Z. *A Survey of Recent Advances in Face Detection* [online]. 2010, : 17 [cit. 2015-12-10].
- [65] ZHANG, S., ZHAO, X., LEI, B., *Facial Expression Recognition Using Sparse Representation*, [online], Taizhou, 2012, v. 11, n. 8, p. 440-452, ISSN: 2224-2678
- [66] ZHAO, W., R. CHELLAPPA, P. J. PHILLIPS a A. ROSENFELD. Face recognition: A Literature Survey. *ACM Computing Surveys* [online]. USA, 2003, 35(4), 399-458 [cit. 2016-05-08]. DOI: 10.1145/954339.954342. ISSN 03600300. Dostupné z: <http://portal.acm.org/citation.cfm?doid=954339.954342>
- [67] ZHU, Xiangxin RAMANAN, Deva. *Face Detection, Pose Estimation, and Landmark Localization in the Wild* [online]. California, USA, 2009, : 8 [cit. 2016-01-03].

SEZNAM ZKRATEK

AAM	Active Appearance Model, aktivní vzhledový model
ARJ	Archived by Robert Jung, komprimační program
BMP	Barevný formát obrazu BitMaP
CCITT	Huffmanovo kódování, kompresní metoda
CIE L^*a^*b	Barevný prostor definovaný světelností, barevnou škálou mezi zelenou a červenou barvou, a mezi modrou a žlutou
CIE L^*u^*v	Barevný prostor definovaný
CIE xyz	Matematicky definovaný barevný prostor vytvořený podle vnímání barev lidí
CMY	Barevný prostor definovaný třemi barevnými kanály: azurová, purpurová, žlutá
CMYK	Barevný prostor CMY doplněný o kanál černé barvy
DCT	Discrete Cosine Transform, diskrétní kosinová transformace
DFFS	Distance from face space, indikátor existence tváře pomocí pozorování globálního minima distanční mapy
EBP	Error Back Propagation, metoda učení neuronové sítě zpětným šířením chyby
EQ	Ekvivalentní ohnisko (35mm)
FA	Faktorová analýza
FLD	Barevná transformace pro nalezení úst
FLDA	Feasibility of Linear Discriminant Analysis
GIF	Graphic Interchange Format, barevný formát obrazu
HMM	Hidden Markov Models, skryté Markovské modely
HSB	Jiné označení pro barevný prostor HSV
HSD	Jiné označení pro barevný prostor HSL
HSI	Jiné označení pro barevný prostor HSL
HSL	Barevný prostor definovaný barevným tónem, sytostí a světlostí
HSV	Barevný prostor definovaný barevným tónem, sytostí a jasovou hodnotou

ICA	Independent Component Analysis, analýza nezávislých komponent
JPEG	Joint Photographic Experts Group, barevný formát obrazu
KNN	K-Nearest-Neighbor, Algoritmus nejbližších sousedů
LDA	Lineární diskriminační analýza
LZ77	Algoritmus bezeztrátové komprese
LZW	Slovníkové kódování, kompresní metoda
MPEG	Moving Picture Experts Group, barevný formát videa
PCA	Principal Component Analysis, analýzy hlavních komponent
PDM	Point Distributed Models, kontura sestavenou z určitého počtu bodů rozmístěných v prostoru
PNG	Portable Network Graphics, barevný formát obrazu
PCX	Barevný formát obrazu
RGB	Barevný prostor definovaný třemi barevnými kanály: červená, zelená, modrá
RGBA	Barevný prostor definovaný třemi barevnými kanály: červená, zelená, modrá a průhledností barev
RLE	Run Length Encoding, kompresní metoda
SVM	Support Vector Machine
TIFF	Tagged Image File Format, barevný formát obrazu
TSL	Barevný prostor definovaný odstínem, saturací a světlostí
WPCA	Váhovaná analýza hlavních komponent
YES	Barevný prostor definovaný světelností, rozdílem mezi zelenou a červenou barvou, a rozdílem mezi modrou a žlutou
YCbCr	Barevný prostor definovaný světelností, rozdílem modré barvy a světelností, a rozdílem červené barvy a světelností
YIQ	Barevný prostor definovaný světelností, barevnou škálou mezi oranžovou a azurovou barvou a barevnou škálou mezi purpurovou a zelenou
YPbPr	Ekvivalent prostoru YCbCr pro analogové signály
YUV	Barevný prostor definovaný světelností, a dvěma barevnými složkami
ZIP	Souborový formát pro kompresi

A VÝSLEDKY KLASIFIKACE VÝRAZŮ POMOCÍ NEURONOVÝCH SÍTÍ

A.1 Výsledky klasifikace výrazů pomocí neuronové sítě naučené na výstupech vlastní banky filtrů

trénovací sada obrazů

rozpoznávaný výraz	1	2	3	4	
1	82 20.5%	0 0.0%	13 3.3%	1 0.3%	85.4% 14.6%
2	1 0.3%	99 24.8%	1 0.3%	0 0.0%	98.0% 2.0%
3	17 4.3%	1 0.3%	86 21.5%	0 0.0%	82.7% 17.3%
4	0 0.0%	0 0.0%	0 0.0%	99 24.8%	100% 0.0%
	82.0% 18.0%	99.0% 1.0%	86.0% 14.0%	99.0% 1.0%	91.5% 8.5%
	1	2	3	4	
	předpokládaný výraz				

sada obrazů 2

rozpoznávaný výraz	1	2	3	4	
1	75 18.8%	11 2.8%	27 6.8%	1 0.3%	65.8% 34.2%
2	1 0.3%	85 21.3%	1 0.3%	4 1.0%	93.4% 6.6%
3	23 5.8%	2 0.5%	71 17.8%	0 0.0%	74.0% 26.0%
4	1 0.3%	2 0.5%	1 0.3%	95 23.8%	98.0% 4.0%
	75.0% 25.0%	86.0% 15.0%	71.0% 29.0%	96.0% 5.0%	81.5% 18.5%
	1	2	3	4	
	předpokládaný výraz				

sada 3

rozpoznávaný výraz	1	2	3	4	
1	72 13.3%	13 2.4%	23 4.3%	8 1.5%	62.1% 37.9%
2	1 0.2%	70 13.0%	4 0.7%	6 1.1%	86.4% 13.6%
3	56 10.4%	30 5.6%	94 17.4%	38 7.0%	43.1% 56.9%
4	6 1.1%	22 4.1%	14 2.6%	83 15.4%	66.4% 33.6%
	53.3% 46.7%	51.9% 48.1%	69.6% 30.4%	61.5% 38.5%	59.1% 40.9%
	1	2	3	4	
	předpokládaný výraz				

sada 4

rozpoznávaný výraz	1	2	3	4	
1	54 11.2%	12 2.5%	28 5.8%	4 0.8%	55.1% 44.9%
2	3 0.6%	52 10.8%	2 0.4%	5 1.0%	83.9% 16.1%
3	59 12.3%	30 7.9%	81 16.8%	39 8.1%	37.3% 62.7%
4	4 0.8%	18 3.7%	9 1.9%	73 15.2%	70.2% 29.8%
	45.0% 55.0%	43.3% 56.7%	67.5% 32.5%	66.3% 33.7%	54.1% 45.9%
	1	2	3	4	
	předpokládaný výraz				

sada 5

rozpoznávaný výraz	1	2	3	4	
1	4 8.6%	0 0.0%	1 2.1%	0 0.0%	90.0% 10.0%
2	3 6.4%	6 12.8%	3 6.4%	0 0.0%	50.0% 50.0%
3	14 29.8%	5 10.6%	8 17.0%	0 0.0%	29.6% 70.4%
4	2 4.3%	1 2.1%	0 0.0%	0 0.0%	0.0% 100%
	17.4% 82.6%	50.0% 50.0%	66.7% 33.3%	NaN% NaN%	38.3% 61.7%
	1	2	3	4	
	předpokládaný výraz				

sada 6

rozpoznávaný výraz	1	2	3	4	
1	1 2.1%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	3 6.3%	7 14.6%	1 2.1%	0 0.0%	63.6% 36.4%
3	16 33.3%	5 10.4%	10 20.8%	0 0.0%	32.3% 67.7%
4	4 8.3%	0 0.0%	1 2.1%	0 0.0%	0.0% 100%
	4.2% 95.8%	68.3% 31.7%	83.3% 16.7%	NaN% NaN%	37.5% 62.5%
	1	2	3	4	
	předpokládaný výraz				

A.2 Výsledky klasifikace výrazů pomocí neuronové sítě naučené na výstupech první banky Gaborových filtrů

trénovací sada obrazů

1	93 23.3%	1 0.3%	14 3.5%	0 0.0%	66.1% 13.9%
2	1 0.3%	95 23.8%	0 0.0%	0 0.0%	99.0% 1.0%
3	6 1.5%	2 0.5%	84 21.0%	0 0.0%	91.3% 8.7%
4	0 0.0%	2 0.5%	2 0.5%	100 25.0%	96.2% 3.8%
	93.0% 7.0%	95.0% 5.0%	84.0% 16.0%	100% 0.0%	93.0% 7.0%
	1	2	3	4	
	předpokládaný výraz				

sada obrazů 2

1	73 18.3%	11 2.8%	27 6.8%	2 0.5%	64.6% 35.4%
2	7 1.8%	72 18.0%	6 1.5%	4 1.0%	80.9% 19.1%
3	17 4.3%	6 1.5%	64 16.0%	3 0.8%	71.1% 28.9%
4	3 0.8%	11 2.8%	3 0.8%	91 22.8%	84.3% 15.7%
	73.0% 27.0%	72.0% 28.0%	64.0% 36.0%	91.0% 9.0%	75.0% 25.0%
	1	2	3	4	
	předpokládaný výraz				

sada 3

1	84 15.6%	67 12.4%	76 14.1%	46 8.5%	30.8% 69.2%
2	13 2.4%	25 4.6%	9 1.7%	9 1.7%	44.6% 55.4%
3	24 4.4%	27 5.0%	41 7.6%	52 9.6%	28.5% 71.5%
4	14 2.6%	16 3.0%	9 1.7%	28 5.2%	41.8% 58.2%
	62.2% 37.8%	18.5% 81.5%	30.4% 69.6%	20.7% 79.3%	33.0% 67.0%
	1	2	3	4	
	předpokládaný výraz				

sada 4

1	68 14.1%	53 11.0%	72 15.0%	39 8.1%	29.3% 70.7%
2	15 3.1%	24 5.0%	10 2.1%	7 1.5%	42.9% 57.1%
3	26 5.4%	21 4.4%	28 5.8%	52 10.8%	22.0% 78.0%
4	11 2.3%	22 4.6%	10 2.1%	23 4.8%	34.8% 65.2%
	56.7% 43.3%	20.0% 80.0%	23.3% 76.7%	19.0% 81.0%	29.7% 70.3%
	1	2	3	4	
	předpokládaný výraz				

sada 5

1	8 17.0%	1 2.1%	5 10.6%	0 0.0%	57.1% 42.9%
2	8 17.0%	8 17.0%	4 8.5%	0 0.0%	40.0% 60.0%
3	5 10.6%	2 4.3%	2 4.3%	0 0.0%	22.2% 77.8%
4	2 4.3%	1 2.1%	1 2.1%	0 0.0%	0.0% 100%
	34.8% 65.2%	66.7% 33.3%	16.7% 83.3%	NaN% NaN%	38.3% 61.7%
	1	2	3	4	
	předpokládaný výraz				

sada 6

1	15 31.3%	9 18.8%	9 18.8%	0 0.0%	45.5% 54.5%
2	3 6.3%	1 2.1%	2 4.2%	0 0.0%	16.7% 83.3%
3	3 6.3%	2 4.2%	1 2.1%	0 0.0%	16.7% 83.3%
4	3 6.3%	0 0.0%	0 0.0%	0 0.0%	0.0% 100%
	62.5% 37.5%	8.3% 91.7%	8.3% 91.7%	NaN% NaN%	35.4% 64.6%
	1	2	3	4	
	předpokládaný výraz				

A.3 Výsledky klasifikace výrazů pomocí neuronové sítě naučené na výstupech druhé banky Gaborových filtrů

trénovací sada obrazů

1	81 20.3%	1 0.3%	13 3.3%	0 0.0%	85.3% 14.7%
2	3 0.8%	91 22.8%	0 0.0%	0 0.0%	96.8% 3.2%
3	15 3.8%	4 1.0%	87 21.8%	2 0.5%	80.6% 19.4%
4	1 0.3%	4 1.0%	0 0.0%	98 24.5%	96.1% 4.9%
	81.0% 19.0%	91.0% 9.0%	87.0% 13.0%	98.0% 2.0%	89.3% 10.8%
	1	2	3	4	

předpokládaný výraz

sada obrazů 2

1	60 15.0%	8 2.0%	32 8.0%	1 0.3%	69.4% 40.6%
2	12 3.0%	77 19.3%	9 2.3%	4 1.0%	75.5% 24.5%
3	22 5.5%	8 2.0%	49 12.3%	3 0.8%	69.8% 40.2%
4	6 1.5%	7 1.8%	10 2.5%	92 23.0%	90.0% 20.0%
	60.0% 40.0%	77.0% 23.0%	49.0% 51.0%	92.0% 8.0%	69.5% 30.5%
	1	2	3	4	

předpokládaný výraz

sada 3

1	36 6.7%	24 4.4%	30 5.6%	23 4.3%	31.9% 68.1%
2	9 1.5%	24 4.4%	10 1.9%	14 2.6%	42.9% 57.1%
3	48 8.9%	37 6.9%	45 8.3%	46 8.5%	25.6% 74.4%
4	43 8.0%	50 9.3%	50 9.3%	52 9.6%	26.7% 73.3%
	26.7% 73.3%	17.8% 82.2%	33.3% 66.7%	38.5% 61.5%	29.1% 70.9%
	1	2	3	4	

předpokládaný výraz

sada 4

1	38 7.9%	24 5.0%	30 6.2%	17 3.5%	34.9% 65.1%
2	3 0.6%	28 5.8%	1 0.2%	8 1.7%	70.0% 30.0%
3	38 7.9%	25 5.2%	46 9.6%	37 7.7%	31.5% 68.5%
4	41 8.5%	43 8.9%	43 8.9%	59 12.3%	31.7% 68.3%
	31.7% 68.3%	23.3% 76.7%	38.3% 61.7%	48.6% 51.2%	35.6% 64.4%
	1	2	3	4	

předpokládaný výraz

sada 5

1	5 10.4%	1 2.1%	1 2.1%	0 0.0%	71.4% 28.6%
2	1 2.1%	3 6.3%	3 6.3%	0 0.0%	42.9% 57.1%
3	3 6.3%	1 2.1%	4 8.3%	0 0.0%	80.0% 20.0%
4	15 31.3%	7 14.6%	4 8.3%	0 0.0%	0.0% 100%
	20.8% 79.2%	25.0% 75.0%	33.3% 66.7%	NaN% NaN%	25.0% 75.0%
	1	2	3	4	

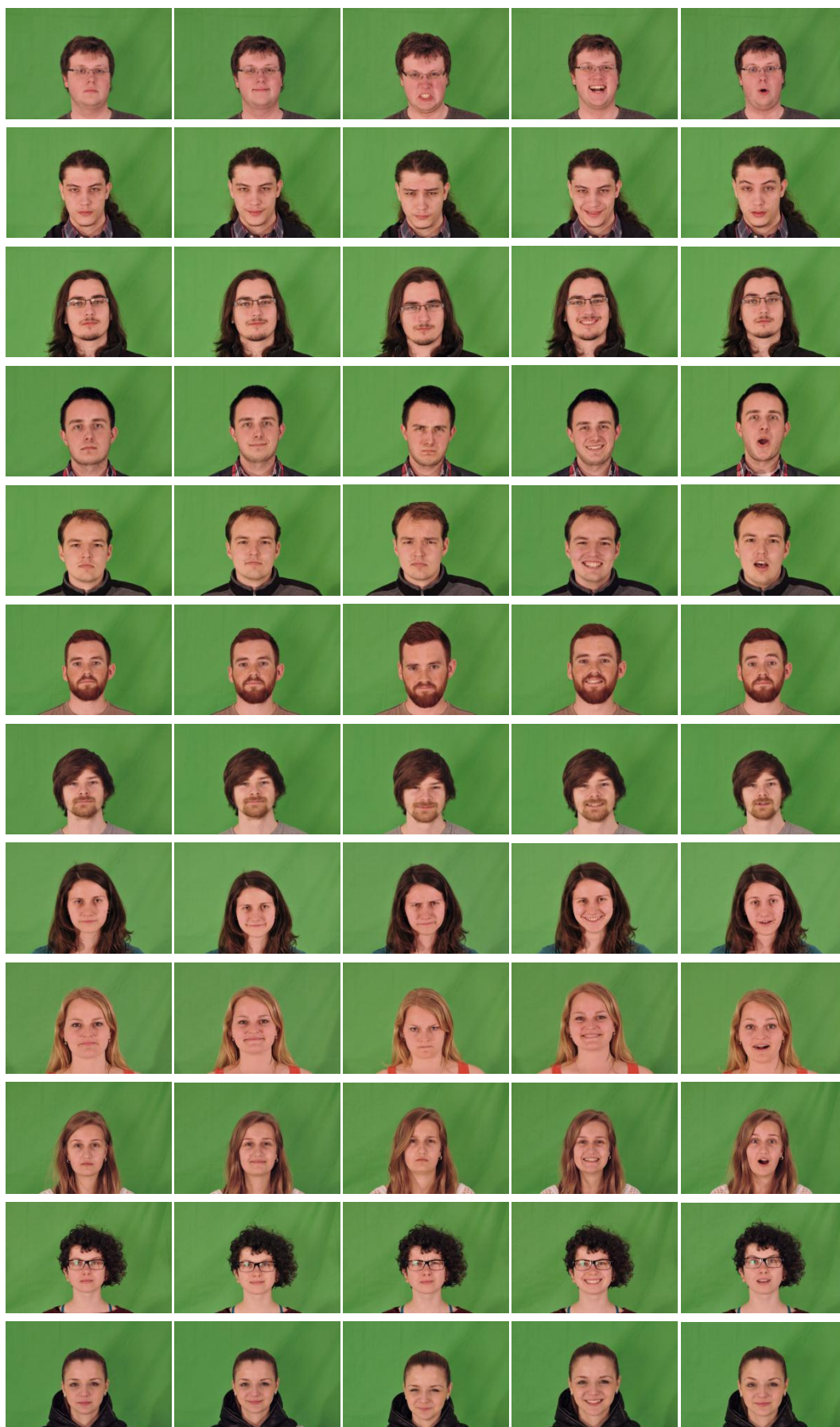
předpokládaný výraz

sada 6

1	3 6.3%	1 2.1%	1 2.1%	0 0.0%	60.0% 40.0%
2	6 12.5%	1 2.1%	2 4.2%	0 0.0%	11.1% 88.9%
3	8 16.7%	5 10.4%	6 12.5%	0 0.0%	31.6% 68.4%
4	7 14.6%	5 10.4%	3 6.3%	0 0.0%	0.0% 100%
	12.5% 87.5%	8.3% 91.7%	50.0% 50.0%	NaN% NaN%	20.8% 79.2%
	1	2	3	4	

předpokládaný výraz

B VLASTNÍ DATABÁZE OBRAZŮ



C SOUPIS SOUBORŮ PŘILOŽENÝCH NA CD

- Text diplomové práce:
- *marketa_vranova_DP.pdf*
- Obrazy vlastní databáze 12 osob s různými výrazy obličeje
- Příklad obrazů z AR databáze použité k učení neuronové sítě
- Zdrojové kódy programu (programováno v prostředí Matlab)
hlavní skript: *rozpoznávání_vyrazu_obliceje.m*
funkce volané hlavním skriptem: *banka1.m*, *banka2.m*, *banka3.m*,
fceHSV.m, *fceRGB.m*, *fceYCC.m*, *fceocinos.m*, *upravy_obrazu.m*
- Uložené proměnné využívané hlavním skriptem - tři naučené neuronové sítě:
bestNet100_20_00322.mat,
bestNetWZ20_00458.mat,
bestNetW100_30_00393.mat